DOCUMENT RESUME

ED 259 582 FL 015 146

AUTHOR	Geis, Michael L., Ed.
TITLE	Studies in Generalized Phrase Structure Grammar. Working Papers in Linguistics #31.
INSTITUTION	Ohio State Univ., Columbus. Dept. of Linguistics.
REPORT NO	OSU-WPL-31
PUB DATE	Jul 85
NOTE	199p.; The paper by Hinrichs was presented at the Annual Meeting of the Linguistic Society of America (1983).
PUB TYPE	Collected Works - General (020) -- Speeches/Conference Papers (150) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Conference Papers; English; German; Language Research; *Linguistic Theory; Masters Theses; *Phrase Structure; *Structural Grammar; *Syntax

ABSTRACT

A group of syntactic studies, primarily concerning English and German, within the framework of generalized phrase structure grammar include: "English Adverb Placement in Generalized Phrase Structure Grammar" (Belinda L. Brodie), concerning the placement of modal, evaluative, temporal, and verb phrase adverbs; "Syntactic Conditions on Two Types of English Cliticization in Generalized Phrase Structure Grammar" (Annette S. Bissantz), on auxiliary reduction and complementizer contraction; "A New Approach to Feature Instantiation in Generalized Phrase Structure Grammar" (Erhard W. Hinrichs), concerning the interaction of feature instantiation principles and linear precedence statements in several European languages; "The Syntax of Conditional Sentences" (Michael L. Geis), presenting the syntactic side of general theory of syntax of English conditionals; and "German Adjective Agreement in Generalized Phrase Structure Grammar" (Arnold M. Zwicky), arguing that such agreement should be treated as a government phenomenon. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED259582

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Dept. of Linguistics
Ohio State University

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The Ohio State University

Working Papers in Linguistics # 31

Studies in
Generalized Phrase Structure Grammar

Edited by Michael L. Geis

Papers by

Annette S. Bissantz, Belinda L. Brodie,
Erhard W. Hinrichs, Michael L. Geis,
and Arnold M. Zwicky

The Ohio State University

Department of Linguistics

204 Cunz Hall
1841 Millikin Road
Columbus, Ohio 43210-1229
U. S. A.

July 1985

BEST COPY AVAILABLE

FL 015 146

OSU WPL # 31

Introduction

The papers in this volume are syntactic studies, primarily concerning English and German, within the framework of Generalized Phrase Structure Grammar. The papers by Bissantz and Brodie represent their master's thesis submitted as part of the requirements for an M.A. degree in the Department of Linguistics at Ohio State University. The paper by Brodie concerns the placement of modal (certainly), evaluative (unfortunately), temporal, and verb phrase adverbs, and that of Bissantz concerns auxiliary reduction (Pita's here) and complementizer contraction (Pita wansta go). Hinrichs' paper, which was read at the annual meeting of the Linguistic Society of America in 1983, concerns the interaction between feature instantiation principles and linear precedence (LP) statements, data being drawn from several European languages. Geis' paper represents the syntactic side of a general theory of the syntax of English conditionals worked out with the pholosopher William Lycan. The paper by Zwicky concerns adjective agreement in German; it is argued that such agreement should be treated as a government phenomenon.

M.L.G.

OSU WPL # 31

Table of Contents

Page

Belinda L. Brodie, "English Adverb Placement in Generalized Phrase Structure Grammar" 1-63

Annette S. Bissantz, "Syntactic Conditions on Two Types of English Cliticization in GPSG" 64-121

Erhard W. Hinrichs, "A New Approach to Feature Instantiation in GPSG" 122-129

Michael L. Geis, "The Syntax of Conditional Sentences" . . 130-159

Arnold M. Zwicky, "German Adjective Agreement in GPSG" . . 160-194

OSU WPL 31 (1985) 1-63

English Adverb Placement in Generalized Phrase Structure Grammar*

Belinda Lea Brodie

1. Introduction

1.1 Objectives

While there have been many works in the last couple of decades dealing with the semantics of English adverbs, few syntactic analyses have been presented. The syntactic analyses which have been proposed have generally been motivated by semantic, rather than syntactic, considerations. The methodological approach taken in these works is to assume or argue for a particular semantic treatment of adverbs, and then provide a syntactic analysis compatible with this semantic treatment. This approach has led, I believe, to incorrect conclusions concerning the placement of adverbs. While I agree with the assumption underlying this approach (i.e that there is a correspondence between syntactic structure and semantic interpretation), I object to the priority given semantic considerations and the disregard for syntactic evidence.

The purpose of this work is to provide an analysis of adverb placement which gives priority to syntactic evidence, and which, moreover, accounts for a much wider range of data than previous analyses. In this work, it is assumed that each syntactic rule is associated with a particular semantic rule. However, it is the syntactic rule which limits the set of possible semantic rules, not vice-versa. I believe that this approach is to be preferred in the case of adverb placement, because there is, at least for some types of adverbs (i.e. evaluative, modal, temporal, and frequency adverbs), evidence which supports a unique syntactic treatment, but no evidence which requires a unique semantic analysis. In these cases it is the syntactic rule which will limit the possible semantic analyses. In cases where there is no evidence to decide between two or more alternate syntactic analyses (i.e. VP adverbs), semantic considerations should, of course, be used to determine the correct syntactic analysis, if possible.

In succeeding sections I will propose analyses of the placement of evaluative, modal, temporal, frequency, and VP adverbs. These analyses are given within the framework of Generalized Phrase Structure Grammar (GPSG). GPSG is a monostratal theory of syntax which is preferable to most other current theories of syntax on two grounds:

1. It is more restrictive, in terms of generative capacity. It has the generative capacity of a context-free phrase-structure grammar, rather than that of a more powerful type of grammar.
2. It is associated with a formal semantics (Montague Grammar) the properties of which are well-defined, and which is restricted by the requirement that the semantics be rule-to-rule.

It is shown that within GPSG it is possible to give a small number of syntactic rules which account for all the data accounted for by previous analyses of adverb positions, as well as a good deal of data not previously considered or accounted for. It is possible to show that these syntactic rules will allow for rule-to-rule translations which will yield the proper semantic results. In this paper, I have been able only to provide the semantic rules and give some indication of how certain aspects of the semantics of adverbs are accounted for by this treatment.

Not only is the analysis to be presented preferable to previous analyses because of its simplicity and scope, but also because this analysis, in conjunction with independently motivated aspects of GPSG, accounts for observations concerning adverb stranding which have not previously been given an adequate treatment. The only previous treatment which is close to being observationally adequate is one in which the notion of "trace" is necessarily referred to in a surface filter constraint (cf. Sag (1978, 1980)). The analysis presented here accounts for adverb stranding data without making reference to traces and is, therefore, consistent with Jacobson's (1982: 207) tentative claim that "no constraint in the grammar can explicitly mention gaps." It is significant that this claim can be maintained with respect to adverb stranding within a framework which is already more restrictive than most other current syntactic theories.

1.2 A brief introduction to GPSG

A GPSG consists of two parts--the actual grammar, which includes the set of phrase-structure rules of the language, and the metagrammar, which consists of rules and principles that characterize the phrase-structure rules and express generalizations between rules.

A phrase-structure rule consists of three parts: a rule number, a syntactic rule, and the semantic rule associated with the syntactic rule. In the PS rule below, for example, the rule number is 2, the syntactic rule is VP -> V VP and the semantic rule is V'(VP').

3. <2, [VP -> V VP], V'(VP')>

Rule numbers are used as subcategorization features on the lexical category node introduced by the rule. Thus, the PS rule above is an abbreviation for the rule in 4.

4. <2, [VP -> $V_{[2]}$ VP], V'(VP')>

The phrase-structure rules are characterized by two types of rules of the metagrammar and by feature instantiation principles. Immediate Dominance (ID) rules express possible immediate dominance relations. The ID rule below, for example, states that A may immediately dominate B, C, and D.

5. A -> B,C,D

The immediate dominance relations expressed by any phrase-structure rule must be identical to immediate dominance relations expressed by one of the ID rules. Linear Precedence (LP) rules express the ordering relationships which must hold between sister nodes. The rule below, for example, states that B must precede C in any phrase-structure rule in which B and C are sisters.

6. B < C

Each phrase-structure rule must be consistent with every LP rule of the metagrammar.

Feature instantiation principles govern the distribution of features. The Head Feature Convention, for example, ensures that the head of a phrase has head features identical to those of its mother. Every phrase-structure rule must be consistent with the Head Feature Convention, and all other principles of feature instantiation.

The metagrammar also includes a type of rule, known as a metarule, which does not characterize phrase-structure rules, but instead expresses implicational relationships between ID rules. The metarule below, for example, states that for every ID rule in the grammar in which A dominates some finite set of category symbols X, there is another rule in which A dominates this same set of symbols and also dominates B.

7. <n, [A -> X], (F')> ==> < [A -> X, B], B'(F')>

This rule also states that the semantic interpretation of the rule on the right will be the result of applying the semantic value of B to the semantic value of F, which is a variable ranging over the interpretations of rules characterized by the syntactic rule on the left of the arrow. The rules related by a metarule will have the same rule numbers.

In earlier versions of GPSG, metarules expressed implicational relationships between PS rules, rather than ID rules. In quoting some earlier works, I will give the rule in terms of PS rules rather than ID rules, but, in every case, the metarule could have just as well been given in terms of ID rules.

I will use the slashing metarule, which Gazdar (1982) defines as follows:

> Let G be the set of basic rules (i.e. the set of rules that a grammar not handling unbounded dependencies would require). For any syntactic category B, there will be some subset of the set of the nonterminal symbols V_N each of which can dominate B according to the rules in G. Let us call this set V_B ($V_B \subset V_N$). Now, for any category B $(B \in V)_N$ we can define a (finite) set of derived rules D (B,G) as follows:
>
> $$D\ (B,G) = \ [a/B \rightarrow o_1 \ldots o_i/B \ldots o_n] : [a \rightarrow o_1 \ldots o_i \ldots o_n] \in G \ \& \ 1 < i < n \ \& \ a,\ o_i \in V_B$$

The slashing metarule is described in this passage as applying to PS rules (basic rules) to allow other PS rules (derived rules). I will sometimes refer to the slashing metarule applying to basic phrase-structure rules to yield derived phrase-structure rules, but, in every case, the slashing metarule could just as easily have applied to ID rules to allow new ID rules. To give an example of the application of the slashing metarule: Given the basic rules in 8,

8. VP -> V VP
VP -> V S
S -> NP VP

the derived rules in 9 will be allowed by the slashing metarule.

9. VP/VP -> V VP/VP
VP/VP -> V S/VP
S/VP -> NP VP/VP

In Gazdar and Pullum (1982) the work done by Gazdar's slashing mechanism is carried out by a feature slash. However, whether the slashing mechanism or the feature slash is used is irrelevant to the analyses I will propose.

It is important to point out that, although in giving rules I have used ADVERB (eg. S -> ADV S) rather than ADVERB PHRASE, replacing ADV in these rules with ADVP would require only a slight revision in the statement of rules. In the rules using ADV, I have made use of the lexical status of ADV to subcategorize adverbs with respect to their sisters so that differences in positions of occurrence could be accounted for. If ADV is replaced by ADVP in the proposed rules, this approach is no longer possible. Instead, we must distinguish various categories of ADVP which dominate different lexical categories of adverbs and give rules for the placement of ADVP, allowing different ADVP categories to occur in different rules.

Finally, it should be noted that, for convenience sake, I have replaced all references to V', N'' etc. with VP, NP etc. In the version of GPSG which I adopt both matrix and embedded VP's are assigned one bar, thus the use of VP is not problematic. My use of S corresponds to V'', the maximal projection of V. In the semantic rules which I give I use the type assignments of Klein and Sag (1982) and follow their convention of not mentioning intentions in the semantic translation; however, when quoting rules, I give the semantic translation as it originally appeared.

FOOTNOTE

*I would like to thank Arnold Zwicky, Mike Geis, and especially David Dowty for their helpful comments and criticism on this work. I am of course solely responsible for all errors.

2. Previous Analyses of Adverb Positions in English

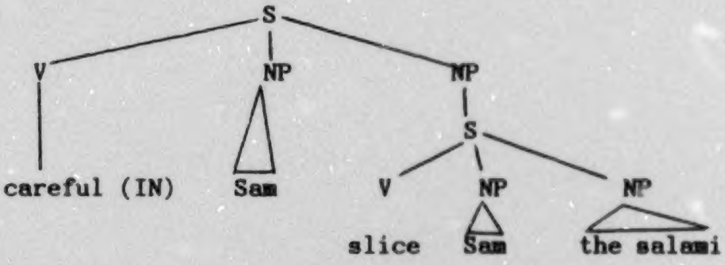
2.1 Lakoff (1965, 1970)

In work by generative semanticists, it was assumed that adverb placement was accounted for by one or more transformational rules. Generative semanticists used the similarities in the selectional restrictions of adverbs and adjectives to argue that adverbs should be transformationally derived from adjectives. Lakoff (1965, 1970) argued that manner adverbs should be derived transformationally from their corresponding adjectives. Similarly, Schreiber (1971) claimed that sentence adverbs should be derived by transformation from their corresponding adjectives.

Lakoff claimed that sentences containing manner adverbs should be derived from the same underlying structures as sentences containing the corresponding adjectives. Both 1 and 2 below were derived from the underlying structure in 3 by means of an adverb lowering transformation. This transformation deletes the occurrence of Sam in the highest clause and moves careful into the lower clause, adding ly onto it.

1. Sam sliced the salami carefully.
2. Sam was careful in slicing the salami.

3.



Lakoff argued for this analysis on two grounds. First of all, he claimed that the elimination of the category Manner Adverb from the set of underlying categories of English would result in simplification of the grammar. The base component is simplified, but the transformational component is complicated by the addition of a rule. Whether or not Lakoff's analysis simplifies the grammar as a whole cannot be determined unless the values of the various elements of the grammar are specified.

Secondly, he claimed that his analysis would eliminate redundancy in the statement of selectional restrictions. In particular, if underlying structures such as 3 are adopted, the anomaly of sentences such as 4 and 5 will be accounted for by the selectional restrictions between underlying subjects and adjectives, and a second set of selectional restrictions between underlying subjects and adverbs need not be included in the grammar.

4. Moss hangs from trees recklessly.
5. Moss is reckless in hanging from trees.

It is, of course, possible to avoid such redundancy without resorting to a transformational derivation of adverbs. In the analysis to be presented here, it will be assumed that meaning postulates (cf.

Dowty (1980)) account for the logical entailment between sentences such as 4 and 5. Once such meaning postulates are adopted, the semantic incoherence of sentences such as 4 will follow from the semantic incoherence of sentences such as 5.

The generative semantics' notion that adverbs are transformationally derived was rejected by transformationalists such as Bowers (1969) and Jackendoff (1972, 1977). They noted that the same arguments which Chomsky (1970) gave in favor of a lexical, rather than transformational, treatment of "derived" nominals also apply in the case of adverbs.

It is now generally accepted that derivational rules, such as the rule creating adverbs from adjectives, apply only in the lexicon. In the analysis to be presented, it will be assumed that a lexical rule (in the sense of Dowty (1978)) derives adverbs from adjectives. It is characteristic of lexical rules that unprincipled exceptions to the rule occur. Schreiber (1971) notes "accidental gaps" such as nicely and improbably, which he considers to be possible but nonoccurring sentential adverbs.

2.2 Jackendoff (1972, 1977)

One of the few syntactic analyses of adverbial positions is that presented in Jackendoff (1972) and slightly revised in Jackendoff (1977). Jackendoff deals with two main classes of adverbs, those traditionally called sentential adverbs and those known as VP adverbs or predicate modifiers.

Jackendoff's analysis is intended to account for the following claimed generalizations concerning the positions of sentential and VP adverbs.

6. A sentential adverb may occupy any position in which it is a daughter of the node S.
7. A VP adverb may occur in any position in which it is a daughter of the node V''.

It is important to note that Jackendoff's notion of V'' is distinct from the notion of V'' used in the version of GPSG adopted here. In Jackendoff's analysis V'' dominates V' and optionally dominates constituents for which the verb in V' is not strictly subcategorized. V' dominates the verb and any constituent for which the verb is strictly subcategorized. Thus, Jackendoff's claimed generalization in 7 predicts that VP adverbs will not intervene between a verb and any constituent for which the verb is strictly subcategorized, since VP adverbs are always daughters of V'', and not V'. Counterexamples to this prediction will be discussed later.

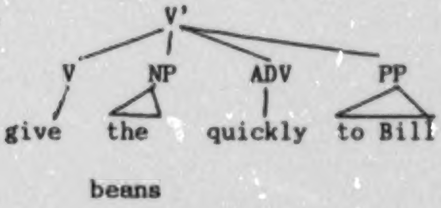
Jackendoff's analysis makes use of the "transportability convention" of Keyser (1968) in order to capture the claimed generalizations in 6 and 7. The transportability convention permits a constituent marked as transportable "to occupy any position in a derived tree so long as the sister relationships with all other nodes in the tree are maintained, that is, as long as it is dominated by the same node." (Jackendoff 1972, p. 67). Jackendoff (1977) claims that sentential adverbs and VP adverbs are transportable constituents. Sentential adverbs will be generated as daughters of S by phrase-structure rules and the transportability convention will allow the sentential adverb to move to any position as long as it remains a daughter of S. VP adverbs will be generated as daughters of V'' by phrase-structure rules and the transportability convention will allow the VP adverb to move to any position as long as it remains a daughter of V''.[1]

It must be pointed out that Jackendoff was not necessarily assuming that these generalizations about adverb positions will hold at the surface structure level. This becomes clear when Jackendoff discusses a class of examples which are problematic for his generalization concerning VP adverbs. The problematic examples are sentences such as 8 in which the VP adverb precedes a PP for which the verb is strictly subcategorized.

8. John gave the beans quickly to Bill.

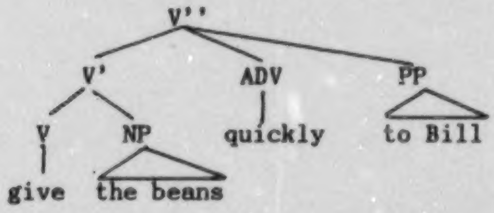
Since give is strictly subcategorized for the PP, the PP will be generated by Jackendoff's phrase-structure rules as a daughter of V'. But Jackendoff's generalization in 7 predicts that the adverb will be a daughter of V'', not V'. Jackendoff considers two solutions. The first solution which he considers is to generate the adverb as a daughter of V'' and then lower it by a transformational rule into position as a daughter of V', yielding the surface structure in 9.

9.



If this solution is adopted, the generalization in 7 is met at the deep structure level, but not at the surface structure level. The other solution which Jackendoff considers is to generate the PP as a daughter of V' but then to raise the PP into position as a daughter of V'', giving the surface structure in 10. On this account, the adverb is a daughter of V'' at both the deep and surface structure levels. Thus, if this solution is adopted, the generalization in 7 is met at both the deep structure and surface structure levels.

10.



Jackendoff does not decide between the two solutions; thus it is not clear whether or not he intends his generalizations to be generalizations about surface structure. If Jackendoff's generalizations concerning positions of sentence and VP adverbs are true generalizations about surface structure positions, then they are easily translated into a monostratal theory in which immediate dominance and linear precedence relations are stated separately (cf. Gazdar and

Pullum (1981)). In immediate dominance (ID) rules, the daughter constituents are unordered with respect to one another. In the ID rule in 11, for example, B, C, and D are unordered. Linear precedence (LP) rules express linear ordering relations between sister constituents. Rule 12, for example, states that B will precede C when they are sisters. The set of phrase structure rules of the grammar includes all rules consistent with some ID rule and every LP rule.

11. A -> B, C, D
12. B < C

In a grammar in which immediate dominance and linear precedence relations are expressed by distinct rules, Keyser's notion of a transportable constituent corresponds to a category which does not appear in any linear precedence rules. Such a category will be unordered with respect to other categories and may, therefore, either precede or follow any of its sister constituents. If Jackendoff's generalizations are meant to hold at the surface structure level, they can be expressed in a version of GPSG which adopts Jackendoff's assumptions about constituent structure by allowing metarules 13 and 14, and by not including the category ADVERB in any LP rules. (I have omitted the semantic translations in 13 and 14, since it is the syntactic generalizations that are at issue here.)

13. <1, V''-> X> ==> <V''-> X ADV>
14. <2, S -> X> ==> <S -> X ADV>

Metarule 13 states that for any ID rule which expands V'' as a finite set of categories X, there will be another ID rule which expands V'' as X plus the category ADVERB. Metarule 14 states that for any ID rule which expands S as X, there will be a rule expanding S as X plus ADVERB. Since ADVERB will not be ordered with respect to any of the categories in X, the metarule in 13 will allow adverbs in the lexical class 1 (i.e. VP adverbs) to appear in any position as daughter of V''. The metarule in 14 will allow adverbs in the lexical class 2 (i.e. S adverbs) to appear in any position as daughter of S.

These two metarules will account for Jackendoff's generalizations in 6 and 7 assuming they refer to surface structure. Unfortunately, this simple analysis cannot be maintained. Jackendoff's generalizations, when considered to apply at the surface structure level, lead to incorrect predictions and rely on unmotivated assumptions about constituent structure.

In order to account for sentences such as 15, for example, Jackendoff must assume that the first auxiliary, but not subsequent ones, is a daughter of S.

15. John will probably leave in the morning.

The only motivation he gives for assuming that the first auxiliary is a daughter of S is that adopting this structure allows the positions of S adverbs in sentences such as 15 to be accounted for by his analysis of adverbs:

> Jackendoff (1972) gives evidence that the first auxiliary is a daughter of S, but that subsequent auxiliaries are not daughters of S. The evidence is that sentence adverbs such as frankly, probably, and evidently occur in all possible positions as daughters of S - initial, final with comma intonation, and before the auxiliary. They also occur after the first auxiliary, but not after subsequent ones.
> (Jackendoff (1977:48)

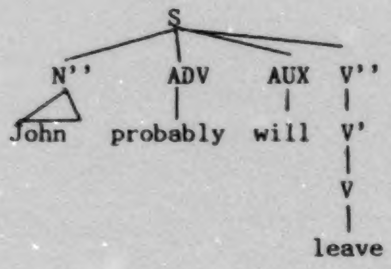
Jackendoff's analysis incorrectly predicts that sentences with S adverbs following the second or third auxiliary should be ungrammatical. But, as Jackendoff (1972) notes, such sentences are not ungrammatical.

16. ?John will have probably been beaten by Bill.
(Jackendoff's example 3.139)

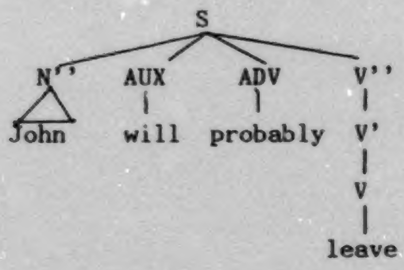
According to Jackendoff's analysis, the surface constituent structures for 17 and 18 would be 19 and 20, respectively.

17. John probably will leave.
18. John will probably leave.

19.

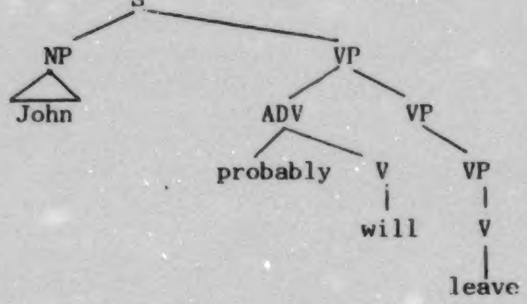


20.

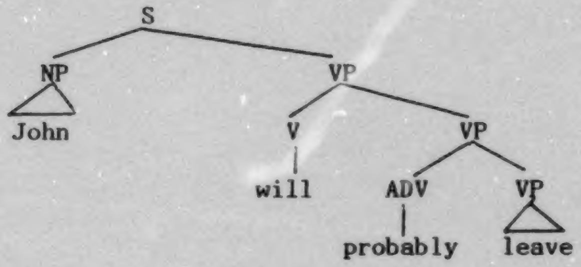


In Section 3, I will argue that the correct constituent structure trees for 17-18 are those in 21 and 22. Since no distinction is made between the matrix VP and embedded VP's in the version of GPSG which I adopt, I use VP, instead of V' or V'', in the trees below.

21.



22.



I will present evidence that sentential adverbs in positions other than clause-initial and clause-final should be accounted for by the phrase-structure rule in 23.

23. VP -> ADV VP

This option was not available to Jackendoff since the phrase-structure rule in 23 does not conform to the rule schema to which, according to the X-bar Convention, all phrase-structure rules must conform. The X-bar Convention requires that one of the daughters in a phrase-structure rule be of the same syntactic category as the mother and one bar level lower than the mother. Thus, the rule in 23 is a counterexample to the X-bar Convention. I will show in chapter 3 that there is ample evidence for the phrase-structure rule in 23, and that, therefore, the X-bar Convention must be rejected.[2]

2.3 Gazdar, Pullum, and Sag (1982)

In Gazdar et al. (1982: 24), the metarule in 24 is given "to handle the facts about sentential adverb placement in the variety of English described by Jackendoff (1972) which only permits the adverb after the first auxiliary verb."

24. <VP -> V VP, F> ==> <VP -> V ADV VP, λP [ADV' (F(P))] >
[+AUX] [-NUL]
[+FIN]

This metarule states that for every rule in the grammar which expands a VP which is marked [+AUXILIARY] and [+FINITE] as V followed by a non-null VP, there will be a rule exactly like this rule except that ADV appears between V and VP. Note that the Head Feature Convention ensures that the V is also [+AUXILIARY] and [+FINITE].

The metarule in 24 is inadequate because it accounts for a very limited range of the positions in which sentential adverbs may occur and also allows for the generation of ungrammatical strings.

Metarule 24 correctly predict that S adverbs occur after the first auxiliary as in 25 and 26.

25. Ed has evidently washed the dishes.
26. Ed will evidently have washed the dishes.

However, as Jackendoff (1972) noted, sentential adverbs may also precede the first auxiliary or main verb. The metarule does not provide for adverbs in these positions. Sentences such as 27-29 are not accounted for by this metarule.

27. Ed obviously has learned French.
28. Ed obviously learned French.
29. Ed obviously will.

The metarule also fails to account for the occurrence of adverbs before the second of two conjoined verbs, as in 30, and for the occurrence of adverbs before the main verbs in sentences in which 'subject-auxiliary inversion' or 'VP fronting' has applied.

30. Ed will catch and probably kill the rabid dog.
31. Will Ed probably kill the rabid dog?
32. ?John said he will definitely pay me and definitely pay me he will.

Gazdar et al. (1982:24) state that "There exists also a less restricted variety in which such adverbs [sentential adverbs] may occur after any auxiliary verb (although the deeper they get in the V', the worse they sound). To handle this variety one needs to delete the [+FIN] specification on the dominant V'." With the [+FIN] specification deleted, the metarule will predict the grammaticality of sentences such as 33.

33. Ed will have evidently washed the dishes.

However, the grammaticality of sentences such as 27-32 is still left unaccounted for.

Gazdar et al. (1982) point out that their metarule predicts the ungrammaticality of strings such as 34 and 35 (their h and i, p. 25) in which the adverb has been stranded.

34. *Kim will obviously.
(with no pause before the adverb)
35. *Kim is obviously
(with no pause before the adverb)

However, this metarule does not predict the ungrammaticality of strings such as 36 and 37.

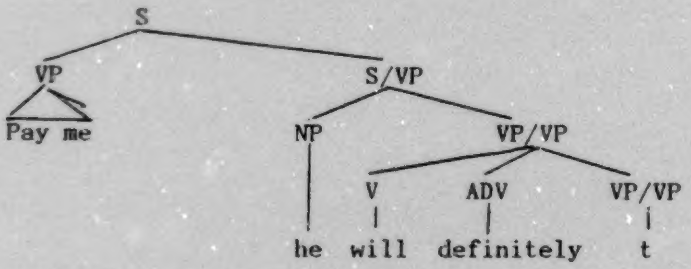
36. *John said he will definitely pay me and pay me he will definitely. (with no pause before the adverb)
37. *I thought John would probably leave and leave he did probably. (with no pause before the adverb)

Metarule 24 in conjunction with the topicalization schema of Gazdar et al. (1982) and the slashing metarule of Gazdar (1981), incorrectly predicts that 36 and 37 are grammatical . The topicalization schema in 38 will allow the rule [S -> VP S/VP].

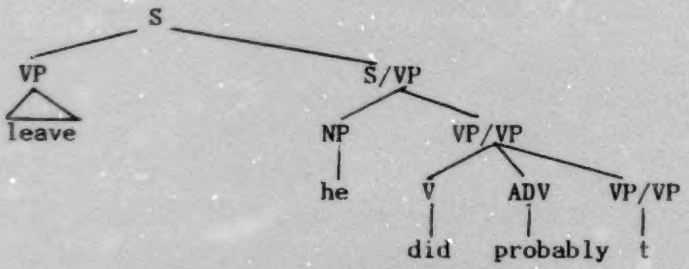
38. <13,[S -> a S/a],λh_a [(S/a)'](a')>

The slashing metarule will apply to the rule on the right of the S-Adverb metarule (i.e. [VP -> V ADV VP]) to give the derived rule [VP/VP -> V ADV VP/VP]. These two rules, along with the derived rule [S/VP -> NP VP/VP] will admit the trees in 40 and 41.

39.


S
VP
Pay me
S/VP
NP
VP/VP
V
ADV
VP/VP
he
will
definitely
t

40.


S
VP
leave
S/VP
NP
VP/VP
he
V
ADV
VP/VP
did
probably
t

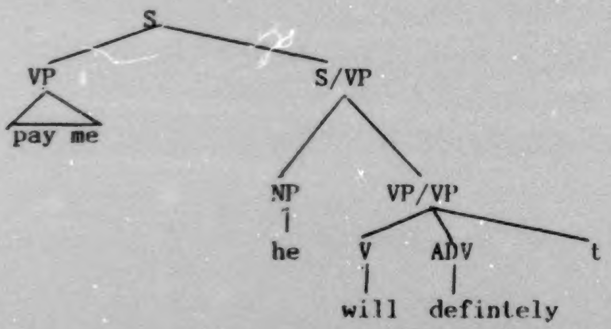
The ungrammatical strings in 36 and 37 will be generated even if the Trace Introduction Metarule (TIM) presented in Sag (1982) is adopted. (The TIM is discussed in more detail in Chapter 6.)

41. Trace Introduction Metarule
[a/B ->...B/B...] ==> [a/B ->...t...]
where a≠B

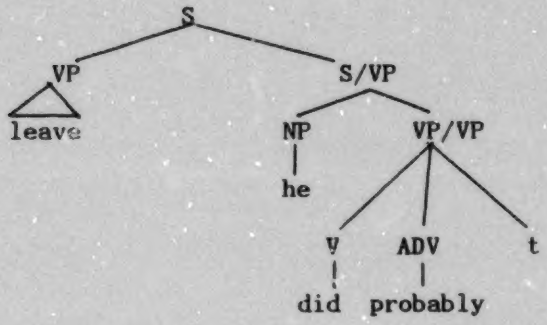
The TIM requires that the node immediately dominating a trace is of the form a/B where a and B are not identical.

If the TIM is adopted, the trees for 36 and 37 will be 43 and 44.

42.


S
VP
pay me
S/VP
NP
VP/VP
he
V
ADV
t
will
defintely

43.



Presumably these trees are admissible, as well as all other trees representing sentences in which 'VP-fronting' has 'applied', because the features on the two VP's of the VP/VP dominating the trace are not the same and therefore a is not equal to B, as is required. If rules of the form [VP/VP ->...VP/VP...] where the VP's of the dominating VP/VP differ in feature specifications are not allowed as input to TIM, then it would no longer be possible to account for 'VP-fronting'.

Even though 36 and 37 will be generated whether or not the TIM is adopted, the ungrammaticality of such sentences could be accounted for if a surface filter, such as the one proposed by Sag (1978, 1980), is employed.

44. $* \left\{ \begin{matrix} \text{Q} \\ \text{ADV} \end{matrix} \right\} \text{t}$

This filter rules out strings in which an adverb (or quantifier) immediately precedes an extraction site. In section 6 arguments will be presented against the surface filter in 44, and it will be shown that, given the analysis proposed in chapter 3, no surface filter is necessary.

In the following section, an analysis within the GPSG framework will be presented which accounts for the data in 27-32 and 36-37, not accounted for by the metarule in 24, as well as other data.

FOOTNOTES

1. In 1972 Jackendoff rejected a transportability analysis of VP adverb positions citing as counterevidence cases of strictly subcategorized adverbs, which only occur in postverbal position. In Jackendoff (1977) it is claimed that strictly subcategorized adverbs are dominated by V' rather than V''. Since only adverbs dominated by S or V'' are subject to the transportability convention, these cases no longer represent counterexamples.
2. Radford (1981:104-106) notes other counterexamples to the X-bar Convention.

3. Evaluative and Modal Adverbs

In this section I will deal with two classes of sentential adverbs which fall into Greenbaum's (1969) category of "attitudinal disjuncts": modal adverbs such as *probably*, *possibly*, *necessarily*, and evaluative adverbs such as *unfortunately*, *luckily*, *preposterously*. Greenbaum does not present any data which would indicate that modal adverbs and evaluative adverbs should be distinguished for syntactic purposes. Jackendoff (1972, 1977) does not distinguish these two classes syntactically. Schreiber (1971:84) claims that "a variety of syntactic arguments can be given...that there are indeed two different types here." The only truly syntactic argument which he gives is that modal adverbs occur in questions, whereas evaluative adverbs do not. In section 3.23, however, it is shown that evaluative adverbs may occur in questions, given the appropriate context. Thus, this purported syntactic difference disappears. In the analysis to be presented modal and evaluative adverbs will be given a uniform syntactic treatment and will belong to the same syntactic class. Such a treatment is possible, because modal and evaluative adverbs are of the same semantic type. When in clause-initial and clause-final positions, they are functions from sentence denotations (i.e. denotations of type $\langle s,t\rangle$) to sentence denotations--they are of the type $\langle \langle s,t\rangle,\langle s,t\rangle\rangle$. When in other positions, it will be claimed, they are functions from VP denotations $\langle\langle s,\langle\langle s,\langle e,t\rangle\rangle,t\rangle\rangle,\langle s,t\rangle\rangle$ to VP denotations--they are of type $\langle\langle\langle s,\langle\langle s,\langle e,t\rangle\rangle,t\rangle\rangle,\langle s,t\rangle\rangle, \langle\langle s,\langle\langle s,\langle e,t\rangle\rangle,t\rangle\rangle,\langle s,t\rangle\rangle\rangle$.[1]

It is useful to consider the adverbs which Schreiber (1971:88) assigns to each category.

1. Modal adverbs: allegedly, certainly, conceivably, evidently, possibly, undoubtedly, unquestionably, clearly, obviously, apparently.
2. Evaluative adverbs: unfortunately, predictably, regrettably, astonishingly, incredibly, interestingly, ironically, luckily, naturally, oddly, predictably, strangely, surprisingly, unbelievably, understandably, unluckily.

What distinguishes these two classes from one another semantically is that the evaluative adverbs are factive, whereas the modal adverbs are not (i.e. *Unfortunately, John left* presupposes, and perhaps entails, that John left, but *Possibly, John left* does not).

The analysis to be presented accounts for the occurrence of modal and evaluative adverbs in sentences in which the adverb has scope over the rest of the sentence. I will not deal with the positioning of adverbs in sentences such as 3-5 in which the adverb does not have scope over the rest of the sentence. In 3 *probably* has scope only over the prepositional phrase *in Westerville*. In 4 probably has scope only over the verb phrase *sing a maudlin song*. In 5 the adverb has scope only over the NP Sharon.

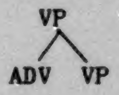
3. We plan to buy a house, probably in Westerville.
4. John will do something for amateur night, probably sing a maudlin song.
5. I gave the book to one of my students, probably Sharon.

I have set aside such sentences from consideration, because I believe that the syntactic analysis of these adverbs will be independent of the syntactic treatment of adverbs in other positions, and thus not immediately relevant to the analyses to be given.

3.1 Evaluative and modal adverbs in positions other than clause-initial and clause-final

In this section it will be argued that modal and evaluative adverbs, when in positions other than clause-initial and clause-final, and when not requiring the intonation pattern required by parentheticals, appear in the configuration in 6.

6.



I will sometimes refer to an adverb in the configuration in 6 as being 'Chomsky-adjoined' to the VP, meaning only that the adverb occurs in this configuration, not that it is actually placed there by a transformation. The lower VP in 6 may dominate either a main or an auxiliary verb.

I will assume that sentences in which the adverb is both preceded and followed by a pause are structurally distinct from sentences in which the adverb is not preceded or followed by a pause. The sentences in 7 and 8, for example, will be assigned distinct structures.

7. John will unfortunately leave.
8. John will, unfortunately, leave.

The adverb in 7 will appear in the configuration in 6, but in 8 it will not. I will assume that the sentence in 8 will have a structure identical to that which a sentence such as 9 has, whatever that may be.

9. John will, as you know, leave.

I am assuming the treatment of auxiliaries given in Gazdar et al. (1982). In this treatment auxiliaries are introduced as daughters of VP by the finite rule schema in 10. The use of features ensures that co-occurence restrictions involving auxiliaries are met. Note that the infinitive marker to is also considered to be a verb.[2]

10. $\langle n, [VP \rightarrow \underset{[+AUX]}{\underset{[a]}{V}}\ \underset{[B]}{VP}], \lambda P[V'(VP'(P))]\rangle$

where values for n, a, and B are given by Table 1.

[n]	a	B	V[n] MEMBERSHIP
[2]	+FIN	+BSE	can, may, must, will etc.
[3]	+FIN	+BSE,-AUX	do
[4]	+ASP	+PSP	have
[5]	+ASP,+COP	+PRP	be
[6]	+COP	+PAS	be
[7]	+INF	+BSE	to
[8]	+FIN,+COP	+INF	is[+COP],ought[-COP]
[9]	+COP	+PRD	be

TABLE 1

My claim is that modal and evaluative adverbs in all of the positions below are 'Chomsky-adjoined' to the following VP.

i. Before a finite main verb:
11. Mark probably left.
12. Mark unfortunately left.

ii. Before a finite auxiliary verb:
13. John probably will leave.
14. John unfortunately will leave.
15. John probably will.

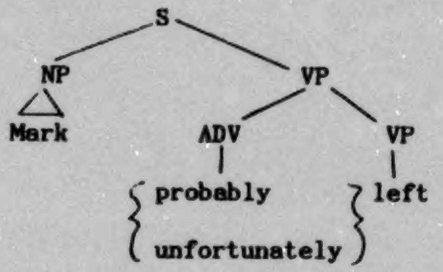
iii. Between a finite auxiliary verb and the main verb:
16. John has probably left.
17. John has unfortunately left.

iv. Between a nonfinite auxiliary verb and the main verb:
18. Ed will have probably washed the dishes by now.
19. Ed will have fortunately washed the dishes by now.

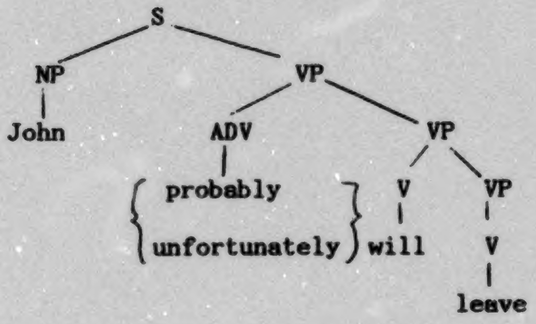
v. Between any two auxiliary verbs.
20. Ed will probably have washed the dishes by now.
21. Ed will fortunately have washed the dishes by now.

The trees for sentences 11-21 are given below.

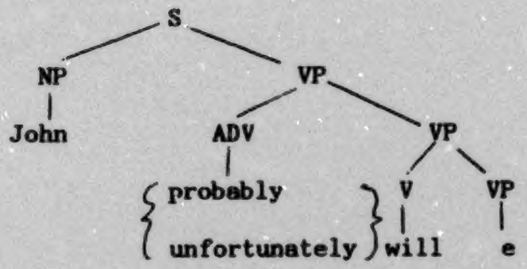
22.



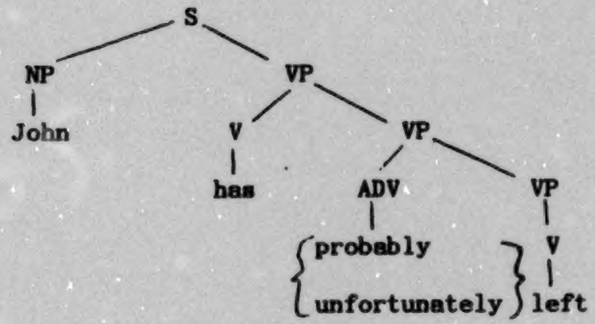
23.


S
NP
VP
John
ADV
VP
probably
unfortunately
V
will
VP
V
leave

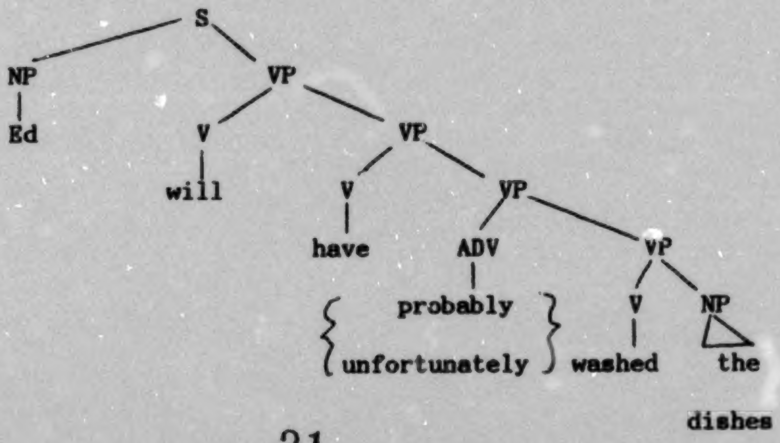
24.


S
NP
VP
John
ADV
VP
probably
unfortunately
V
will
VP
e

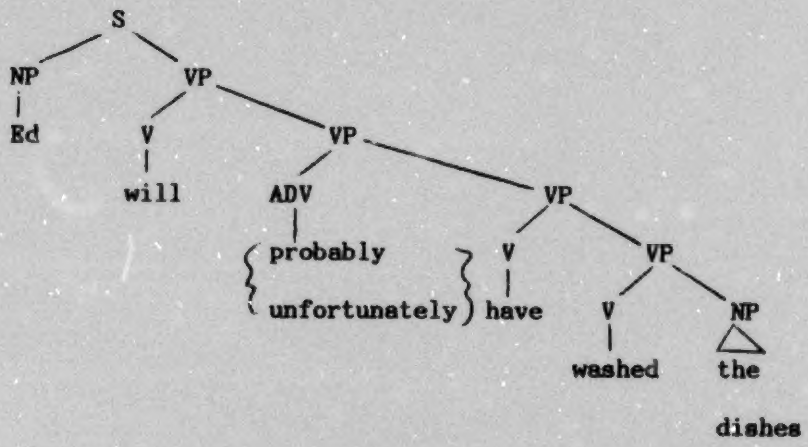
25.


S
NP
VP
John
V
has
VP
ADV
probably
unfortunately
VP
V
left

26.


S
NP
Ed
VP
V
will
VP
V
have
VP
ADV
probably
unfortunately
VP
V
washed
NP
the
dishes

27.



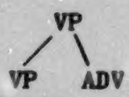
All of the sentences above can be accounted for by adding the ID rule in 28 to the grammar.

28. <1, [VP -> ADV, VP], ADV'(VP')>
[1]
where ADV(1) = the modal and evaluative adverbs

Note that the daughter VP in 28 is the head of the mother VP. Therefore, the daughter VP will have all of the same head features as its mother, because the Head Feature Convention ensures that the head of a phrase will have head features identical to those of its mother node. The ADV will be marked with the rule number (1), because the lexical category introduced by a rule is marked with the rule number. In the semantic rule in 28 the value of the adverb is a function which takes VP type denotations as arguments and yields VP type denotations.

It is necessary to ensure that adverbs of lexical category 1 may not follow their sister VP's. Otherwise, the subtree in 29 will be generated and it will be incorrectly predicted that sentences such as 30 are grammatical.

29.



30. *Patrick went to the bank probably and withdrew money from from his checking account definitely.

We could add the LP rule in 31 to ensure that an adverb of the lexical category 1 must precede a sister VP. However, if we assume, as do Gazdar and Pullum (1982), that English includes a general LP rule requiring lexical categories to precede non-lexical categories, then 30 is not necessary.

31. ADV <VP
[1]

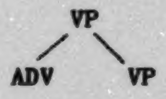
Given the ID rule in 28 and the LP rule in 31 the basic phrase-structure rule in 32 below will be part of the grammar, but the rule in 33 will not.

32. S-Adverb basic rule:
<1, [VP -> ADV VP], ADV' (VP')>
where ADV(1)= the evaluative and modal adverbs
33. <1, [VP -> VP ADV], ADV' (VP')>

I will label 32 the S-Adverb basic rule, since evaluative and modal adverbs have traditionally been known as sentential adverbs, but I do not mean to imply that every adverb which has been labeled a sentential adverb should be introduced by a rule like that in 28.[3]

In the following section I will show that once the rule in 32 is adopted, interaction with a number of independently-motivated rules of GPSG accounts for a wide range of data. First, however, I will use some of this data as basically theory-neutral evidence for the configuration in 6, repeated below.

34.



The examples in 35-37 provide evidence for the higher VP node in 34.

35. John said he would definitely pay me and definitely pay me, he will. [VP Preposing]
36. Two plus two will necessarily equal four and one plus three will, too. [VP Deletion]
37. John probably will swing and possibly will hit the ball. [VP Conjunction]

If 35-37 are indeed examples of VP Preposing, VP Deletion, and VP Conjunction, as they certainly appear to be, then the adverb and following verb phrase must be dominated by VP.

Examples such as 38 provide evidence for the lower VP node in 34.

38. Rhonda has probably been to Dinosaur Park and Jimmy definitely has.

Because of the presence of definitely in the second conjunct, the only interpretation for this sentence is one in which only been to Dinosaur Park has been 'deleted', and not probably been to Dinosaur Park. The semantic rule associated with VP Deletion ensures that the value of a previous VP is eventually plugged in to the translation of the right conjunct. Since the value of been to Dinosaur Park is plugged in to the translation of the right conjunct in 38, been to Dinosaur Park in the left conjunct must be a VP. A semantic analysis of VP Deletion must allow either the value of the lower VP or the value of the higher VP to be plugged in to the translation of the right conjunct. In 36, the most natural reading is one in which the value of the higher VP is plugged in. In 38, the reading in which the value of the lower VP has been plugged in is forced by the presence of definitely.

3.2 Evidence from rule interactions

In this section I will show that once the S-Adverb basic rule in 32 is adopted, interaction with a number of independently-motivated rules of GPSG account for a wide range of data.

3.21 Coordination

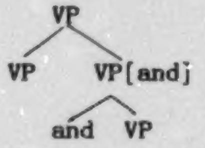
Gazdar (1981:158) proposes the following rule schemata to account for constituent coordination.

39. $\langle 2, [a \rightarrow a_1 \ldots, \underset{[B]}{a_n}], B'(a_1', \ldots, a_n') \rangle$
where $B \in$ [and, or] and a is any syntactic category

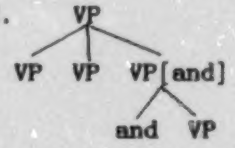
40. $\langle 3, [\underset{[B]}{a} \rightarrow B\ a], a' \rangle$
where $B \in$ [and, or] and a is any syntactic category

If a is VP and B is and, then the schemata in 39 and 40 will produce structures such as 41 and 42.

41.



42.

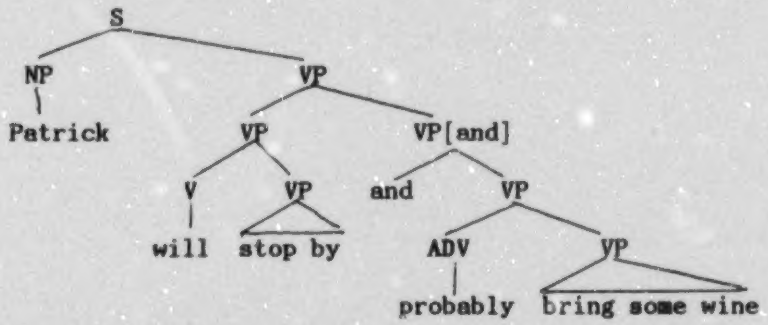


These rule schemata together with the S-Adverb basic rule proposed in section 3.1 predict the grammaticality of sentences such as 43-46 in which adverbs precede verb phrase conjuncts.

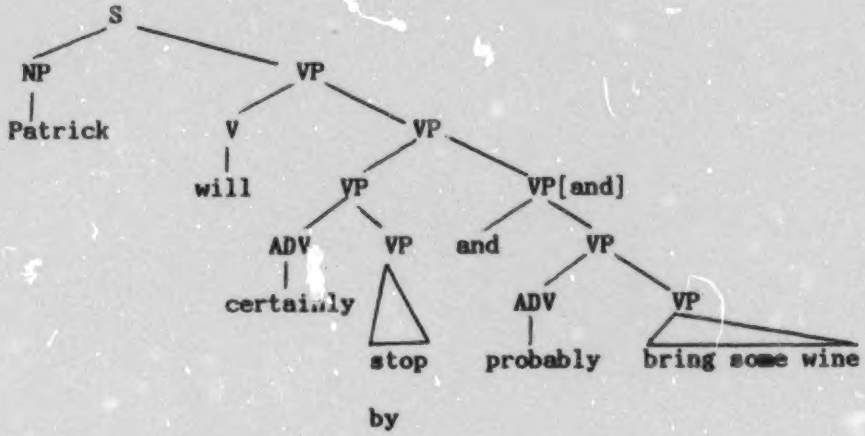
43. Patrick will stop by and probably bring some wine.
44. Patrick will certainly stop by and probably bring some wine.
45. Patrick studied, but probably flunked the test anyway.
46. Patrick probably works hard and definitely enjoys his work.

These sentences will be assigned the following trees:

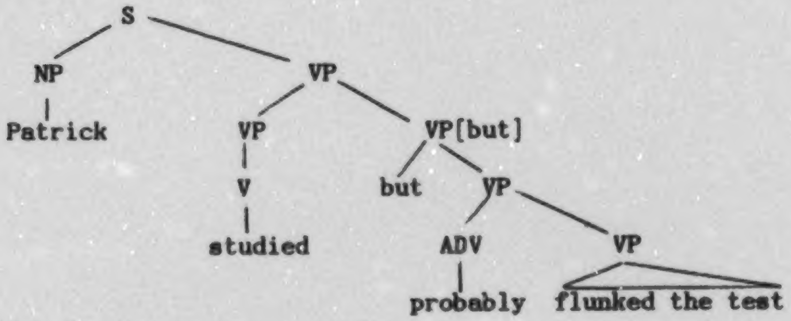
47.



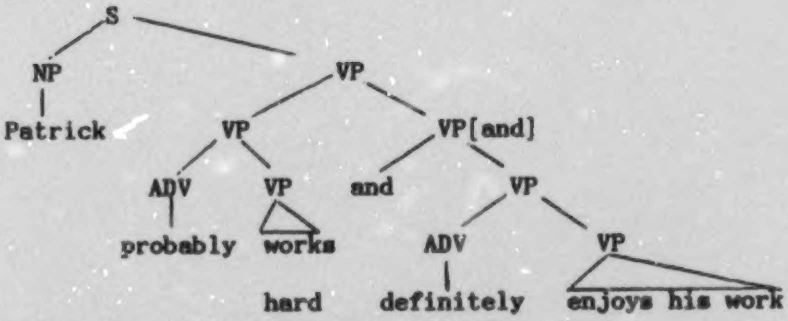
48.


S
NP
VP
Patrick
V
VP
will
VP
VP[and]
ADV
VP
and
VP
certainly
ADV
VP
stop
by
probably
bring some wine

49.


S
NP
VP
Patrick
VP
VP[but]
V
but
VP
studied
ADV
VP
probably
flunked the test

50.


S
NP
VP
Patrick
VP
VP[and]
ADV
VP
and
VP
probably
works
hard
ADV
VP
definitely
enjoys his work

The grammaticality of sentences such as 51-54, with evaluative adverbs, is, of course, also predicted.

51. Patrick will stop by and unfortunately stay for dinner.
52. Patrick will fortunately stop by, but unfortunately stay for dinner.
53. Patrick stopped by and unfortunately stayed for dinner.
54. Patrick fortunately stopped by, but unfortunately stayed for dinner.

Examples such as 43-46 and 51-54 are problematic for Jackendoff's analysis. They are obvious counterexamples to the claim that S-adverbs always occur as daughters of S at the surface structure level. Once Conjunction Reduction is given up, such examples are also counterexamples to the claim that S-adverbs are always daughters of S at the deep structure level.

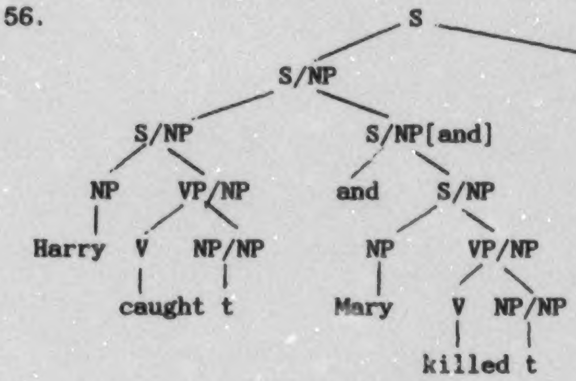
3.22 Right Node Raising

The next set of examples include sentences which have traditionally been described as having undergone the transformation of right node raising. In Gazdar (1981) right node raising structures are accounted for by the rightward displacement schema in 55.

55. <9, [a -> a/B B], λh_B[(a/B)'] (B')>
where a ranges over clausal categories and B can be any phrasal or clausal category.

The rightward displacement schema, the coordination schemata, and the slashing metarule interact to produce structures such as that in Gazdar's example in 56.

56.



The rightward displacement schema permits the rule [S -> S/NP NP].

The coordination schemata permit the rules [S/NP -> S/NP S/NP] and [S/NP -> and S/NP].

The slashing metarule permits the rules [S/NP -> NP VP/NP] and [VP/NP -> V NP/NP].

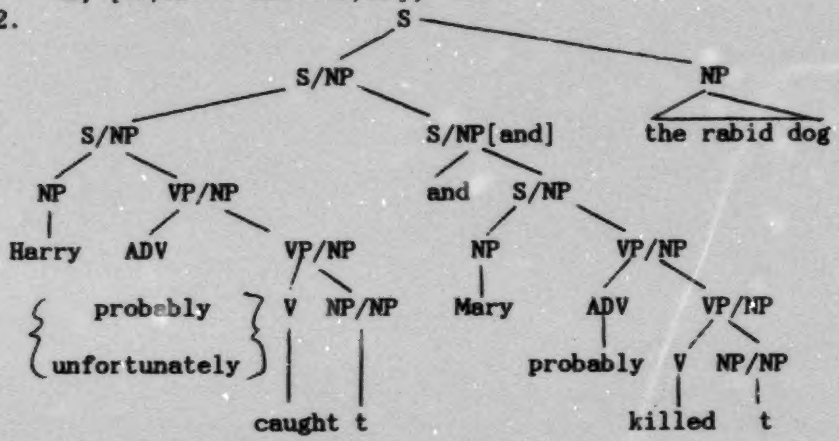
With the addition of the S-Adverb basic rule to the grammar, the grammaticality of right node raising sentences such as the following in which an S-adverb precedes the verb(s) is predicted.

57. Harry probably caught and Mary certainly killed the rabid dog.
58. Harry caught and Mary probably killed the rabid dog.
59. Harry caught and Mary unfortunately killed the rabid dog.
60. Harry fortunately caught and Mary fortunately killed the rabid dog.

The slashing metarule applies to the S-Adverb basic rule, as in 61, and allows the derived rule on the righthand side of the arrow. This derived rule allows the adverb to precede a verb which has had its object raised, as shown in 62.

61. <1, [VP -> ADV VP], F> ==>
    <1, [VP/NP -> ADV VP/NP], F>

62.

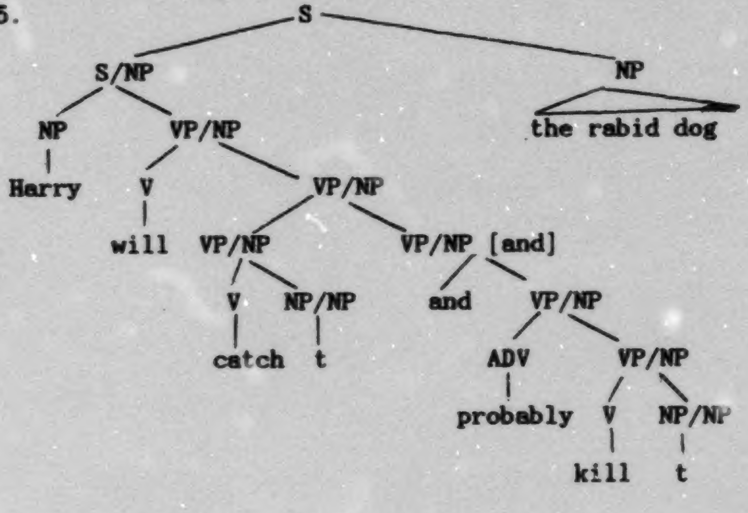

S
S/NP
NP
S/NP
S/NP[and]
the rabid dog
NP
VP/NP
and
S/NP
Harry
ADV
VP/NP
NP
VP/NP
probably
unfortunately
V
NP/NP
Mary
ADV
VP/NP
probably
V
NP/NP
caught
t
killed
t

The rightward displacement schema, the coordination schemata, the slashing metarule, and the derived rule in 61, together, predict that sentences such as the following are grammatical.

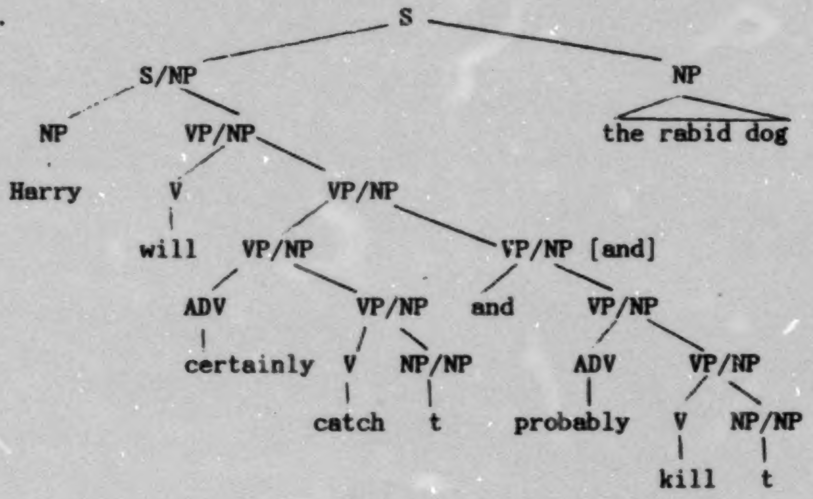
63. Harry will catch and probably kill the rabid dog.
64. Harry will certainly catch and probably kill the rabid dog.

The grammar will assign these sentences the following trees.

65.


S
S/NP
NP
NP
VP/NP
the rabid dog
Harry
V
VP/NP
will
VP/NP
VP/NP [and]
V
NP/NP
and
VP/NP
catch
t
ADV
VP/NP
probably
V
NP/NP
kill
t

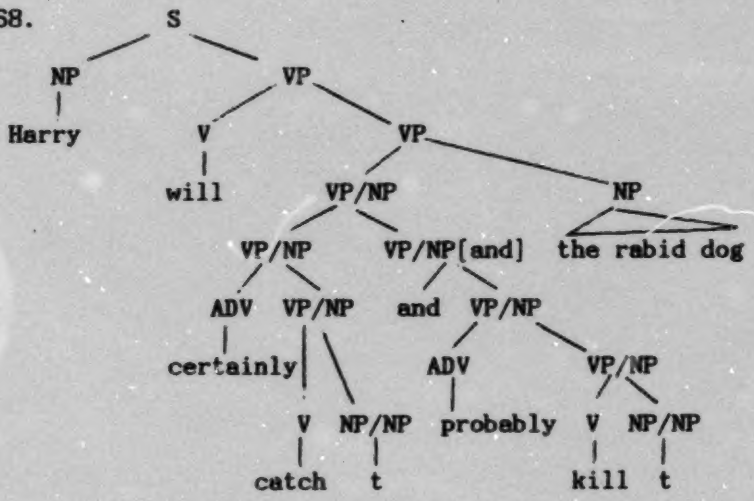
66.


S
S/NP
NP
NP
Harry
VP/NP
the rabid dog
V
VP/NP
will
VP/NP
VP/NP [and]
ADV
VP/NP
and
VP/NP
certainly
V
NP/NP
ADV
VP/NP
catch
t
probably
V
NP/NP
kill
t

The trees in 65 and 66 are right node raising structures, but sentences such as 63 and 64 can be produced without the intonation pattern characteristic of right node raising. Thus, it seems necessary to provide trees for such sentences which do not have a right node raising configuration. I will argue that the grammar should include the rule of "minor right node raising" in 67.[4] Once this rule is adopted, sentences such as 63 and 64 will be assigned two distinct tree structures, one like those in 65 and 66 and another like that in 68.

67. <57, [VP -> VP/NP NP], VP/NP'(NP')>

68.

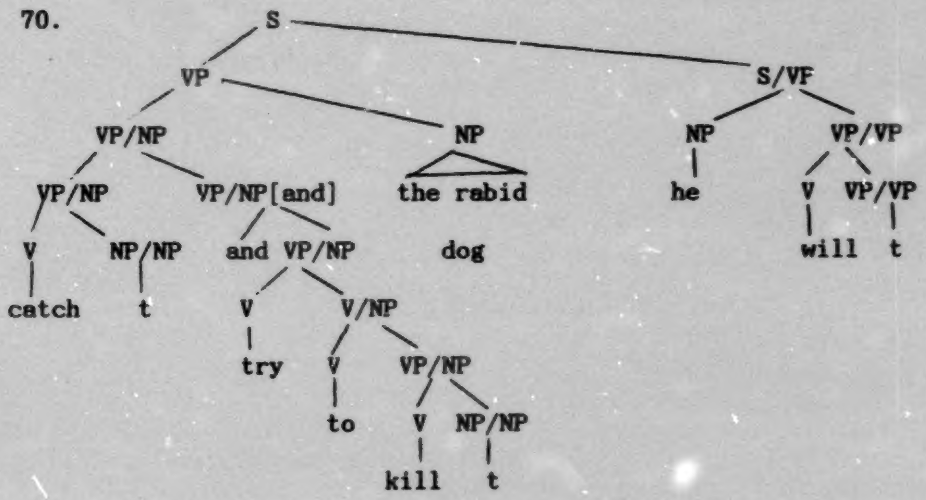

S
NP
VP
Harry
V
VP
will
VP/NP
NP
VP/NP
VP/NP[and]
the rabid dog
ADV
VP/NP
and
VP/NP
certainly
ADV
VP/NP
V
NP/NP
probably
V
NP/NP
catch
t
kill
t

Inclusion of the rule [VP -> VP/NP NP] in the grammar can be motivated by considering sentences such as 69.

69. Ed said he would catch and try to kill the rabid dog and catch and try to kill the rabid dog he will.[5]

If the second conjunct of sentence 69 is to be treated as an example of 'VP-fronting', then the grammar must provide a structural description of catch and try to kill the rabid dog in which the entire phrase is a VP. The minor right node raising rule provides for such a description, as shown in 70.

70.

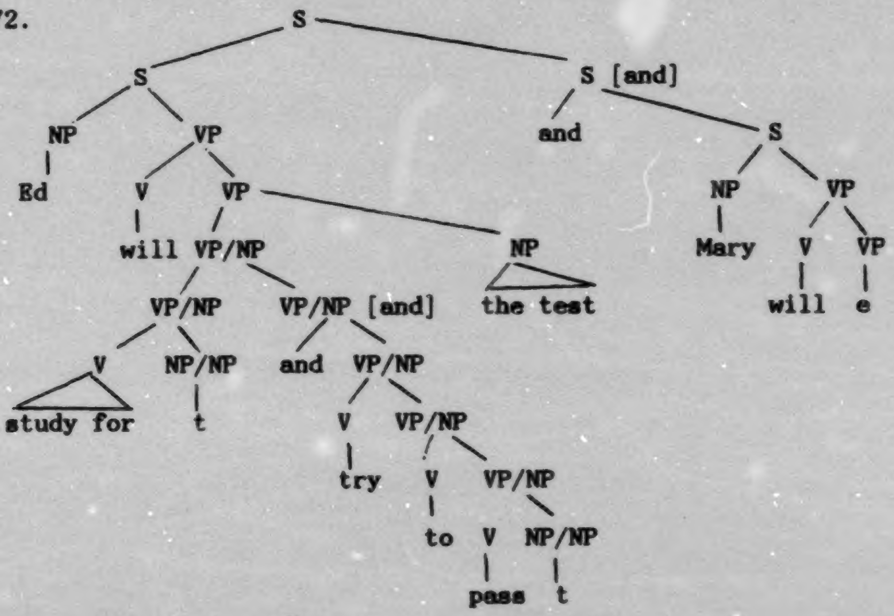

S
VP
S/VP
VP/NP
NP
NP
VP/VP
VP/NP
VP/NP[and]
the rabid
dog
he
V
VP/VP
will
t
V
NP/NP
catch
t
and
VP/NP
V
try
V/NP
V
to
VP/NP
V
kill
NP/NP
t

Sentences such as 71 provide further evidence for minor right node raising.

71. Ed will study for and try to pass the test and Mary will too.

If the second conjunct in sentence 71 is to be treated as an example of 'VP Deletion', then the grammar must allow study for and try to pass the test to be a VP. The rule [VP -> VP/NP NP] will provide a representation in which it is a VP, as in 72.

72.


S
S
S [and]
NP
Ed
VP
V
will
VP
VP/NP
NP
the test
VP/NP
V
study for
NP/NP
t
VP/NP [and]
and
VP/NP
V
try
VP/NP
V
to
VP/NP
V
pass
NP/NP
t
and
S
NP
Mary
VP
V
will
VP
e

It must be noted that inclusion of the rule [VP -> VP/NP NP] will allow for the generation of the following ungrammatical sentences:

73. *John gave a vase the woman who he's dating.
74. *John persuaded that Harold left the woman who's standing over there.
75. *John expected to win the runner from Australia.
76. *John said he would give the woman he's been dating a vase and give a vase the woman he's been dating, he will.

Sentences such as 73-75 will be generated by the grammar anyway because of the rightward displacement rule. One solution which comes to mind for preventing both the rightward displacement rule and the minor right node raising rule from generating these sentences is to simply disallow the following rules from the grammar:

```
77.  VP/NP -> V  NP/NP  NP
78.  VP/NP -> V  NP/NP  S
79.  VP/NP -> V  NP/NP  VP
```

This, however, is not a viable solution because rules 77-79 are needed to generate grammatical sentences such as the following, in which topicalization has applied.

80. Janet, John gave a vase.
81. Kim, John persuaded that Fido runs.
82. Jimmy Carter, John wanted to win.

Unfortunately, I do not have a solution to offer at this time. It should be noted, however, that examples such as 73-75 have been problematic for all previous analyses of right node raising, and that it is very likely that an account of the ungrammaticality of these sentences when assigned right node raising structures will also account for their ungrammaticality when assigned minor right node raising structures.

3.23 Subject-Auxiliary Inversion

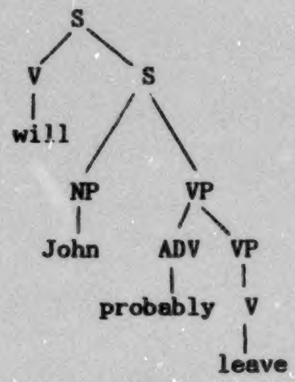
Gazdar et al. (1982) give the following metarule to account for sentences in which 'Subject-Auxiliary Inversion' has applied.

```
83. <[VP -> V  VP], λP[V'(^V'(P))] ==>
    [+FIN]    [α]
    [+AUX]
    <[S -> V  S], V'(^S')>
     [+INV]  [α]
```

The S-Adverb basic rule interacts with this metarule to predict the grammaticality of questions such as 84 and 85. The question in 84 will be assigned the tree in 86.

84. Will John probably leave?
85. Did John fortunately leave?

86.



Both Jackendoff (1972) and Bellert (1977) have argued that questions such as 84 with modal adverbs are syntactically well-formed, but semantically or pragmatically odd.

Jackendoff (1972:84) notes that "many S adverbs do not feel comfortable in questions" and 'stars' 87 (his 3.160) to indicate that it is unacceptable.

87. *Did Frank probably beat all his opponents?

He argues that "purely syntactic approaches to the unacceptability of sentences such as 87 miss the point" and that "a more semantically based analysis is called for, in which there is a reason for these facts" (p. 84).

Bellert (1977:344) states that "Modal sentential adverbs are predicates of the truth: they qualify the truth of the proposition expressed in the same sentence, and they do not qualify it negatively. Neither do they occur in questions" and cites 88 (her 21).

88. $*\left\{\begin{matrix}\text{Has}\\ \text{Will}\end{matrix}\right\}$ John $\left\{\begin{matrix}\text{probably}\\ \text{certainly}\\ \text{evidently}\end{matrix}\right\}$ come?

Such questions are unacceptable, according to Bellert, because "we do not ask questions and at the same time evaluate the truth, or degree of truth, of the proposition that is being questioned" (p. 344).

The explanation offered by Bellert is supported by the observation that, in contexts in which these constraints do not, in general, hold, questions such as those in 87 and 88 are acceptable. A context in which we expect questions to be asked which "at the same time evaluate the truth, or degree of truth, of the proposition that is being questioned" and which "assert a proposition in one and the same sentence" is the courtroom context. Questions such as 89-90 are certainly acceptable in a courtroom setting.

89. In your opinion, has the defendent possibly perjured himself?
90. In your opinion, did John Jones probably commit suicide?

It is clear from such examples that the unacceptability of sentences such as 87 and those in 88 should be accounted for by constraints such as those offered by Bellert, rather than by syntactic constraints. A similar constraint is obviously responsible for the unacceptability of questions such as 85: we do not question a proposition and at the same time comment on the proposition we are questioning. This constraint is violated in a courtroom setting, perhaps because lawyers are often obviously assuming the truth of a proposition while at the same time asking for a witness to evaluate the truth of the proposition. So, in 91, for example, the questioner is commenting on the proposition, while at the same time asking that the respondent evaluate its truth.

91. Did this woman unfortunately get involved in a life of crime?

I will conclude that the questions in 84-85 and 87-88 are syntactically well-formed and that the S-Adverb basic rule and 'Subject-Auxiliary Inversion' metarule interact to yield correct predictions.

3.24 VP Fronting

Gazdar et al. (1982) account for 'VP-fronting' by the more general rule of Topicalization, repeated below. They claim that "the phenomenon commonly referred to as 'VP-fronting' is simply a special case of topicalization and can therefore be subsumed under schema 13 [92 below] by allowing a to range over frontable V' [VP] types"(p. 18).

92. <13, [S -> a S/a],λh_a[((S/a)')(a')]>
when a = VP, then a is to be [-FIN, -INF, -ASP]

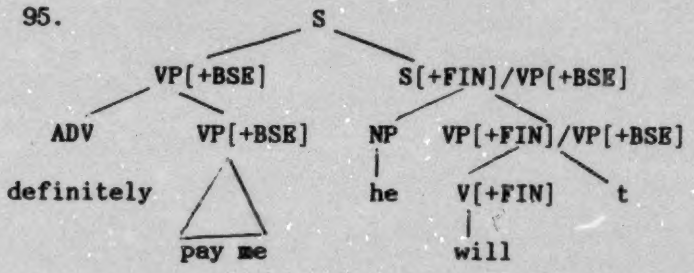
The S-Adverb basic rule and the topicalization schema in 92 correctly predict that sentences such as 93 and 94 are grammatical.

93. ?John said he will definitely pay me and definitely pay me he will.
94. ??I thought John would probably leave to avoid seeing his mother and probably leave he did.

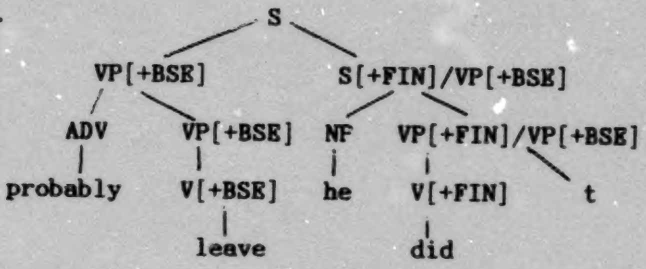
While sentences such as 93 and 94 may not be fully acceptable, they are certainly not ungrammatical. It seems likely that semantic and/or pragmatic constraints on VP Topicalization, similar to those discussed by Prince and Prince (1980) for NP Topicalization, may be responsible for the oddness of such sentences.

These sentences will be assigned the structures in 95 and 96. The topicalization schema allows the rule [S -> VP S/VP] and the S-Adverb basic rule allows the rule [VP -> ADV VP].

95.



96.



It is also predicted that sentences such as 97 are grammatical, since evaluative adverbs are also introduced by the S-Adverb basic rule.

97. Bob knew Bill would unfortunately flunk the test, and
unfortunately flunk the test, he will.

However, sentences such as 97 are much worse than sentences such as 93 and 94. But the differences in acceptability are expected, given the differences in acceptability between 98 and 99.

98. Bob knew Bill would probably flunk the test.
99. Bob knew Bill would unfortunately flunk the test.

Sentence 98 is fine, but 99 is questionable. The oddness of sentences such as 99 obviously has to do with the function of evaluative adverbs: they are expressions of the speaker's attitude. Parentheticals such as I think also express the speaker's attitude. Although parentheticals which express the speaker's attitude occur freely between the auxiliary and VP in main clauses, as in 100, they do not occur in this position in embedded clauses.

100. Bill will, I think, flunk the test.
101. Bob knew Bill will, I think, flunk the test.

The oddness of sentences such as 99 should be explained in part by the same pragmatic or semantic constraint which explains the oddness of sentences such as 101.

Given the rule of Topicalization and the derived rule in 102, it would seem that ungrammatical strings such as that in 103, in which the adverb is stranded, will result.

102. VP/VP -> ADV VP/VP
103. *John said he would definitely pay me, and
pay me he will definitely. (with no pause before the adverb)

However, once the Trace Introduction Metarule is adopted, strings such as that in 103 will not be produced. I will not explain here why this is the case, since a detailed explanation is given in section 6.

3.25 VP Deletion

Gazdar et al. (1982) give the following metarule to account for 'VP Deletion':

```
104.  VPD:   < [VP -> V  VP], F > ==> <[VP -> V  VP], F>
                [+AUX]                          [+NUL]
                [-PRP]
                [-GER]
```

The metarule in 104 "takes any V' [VP] [+AUX, -PRP, -GER] rule which expands as V followed by V' [VP], and simply adds the feature +NUL to the complement V' [VP]" (p. 606). The rule in 105 introduces e, which represents the empty string.

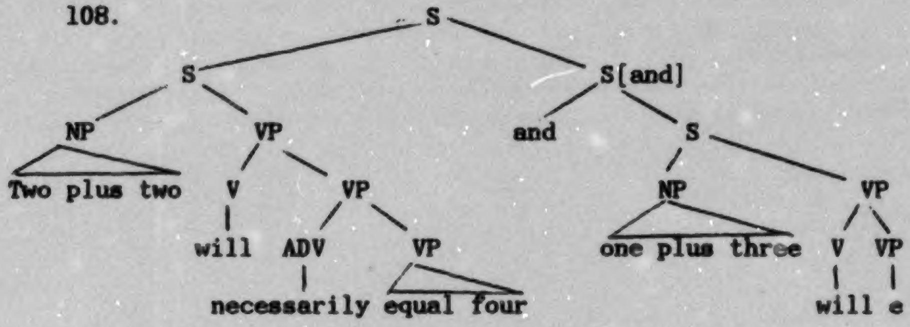
```
105.  <[VP -> e, v >
      [+NUL]
      where v is a contextually bound variable ranging over
      VP denotations.
```

Given this analysis of VP Deletion and the S-Adverb basic rule, sentences such as 106 and 107 will be assigned the trees in 108 and 109, respectively.

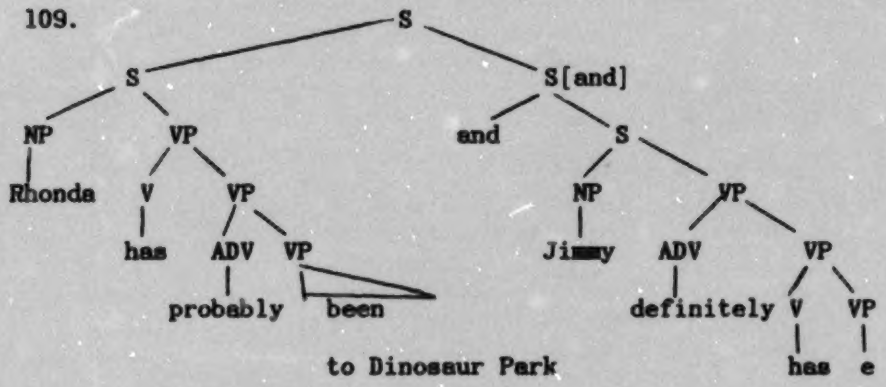
106. Two plus two will necessarily equal four and
one plus three will, too.

107. Rhonda has probably been to Dinosaur Park and
Jimmy definitely has.

108.


S
S
S[and]
NP
VP
and
S
Two plus two
V
VP
NP
VP
will
ADV
VP
one plus three
V
VP
necessarily
equal four
will
e

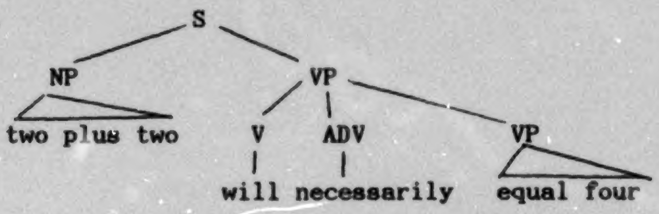
109.


S
S
S[and]
NP
VP
and
S
Rhonda
V
VP
NP
VP
has
ADV
VP
Jimmy
ADV
VP
probably
been
definitely
V
VP
to Dinosaur Park
has
e

The most natural reading of 106 is the one in which the contextually bound variable v takes as its value the denotation of the VP necessarily equal four. In 107, the reading in which v takes as its value the denotation of the VP been to Dinosaur Park is forced by the presence of definitely in the second conjunct.

It should be noted that given the metarule analysis of sentential adverbs of Gazdar et al. (1982) it would be difficult to explain the interpretation of 106 in which what has been 'deleted' semantically in the second conjunct is necessarily equal four, since, according to this analysis, necessarily equal four would not be a constituent. Under their analysis, 106 would be assigned the tree in 110.

110.



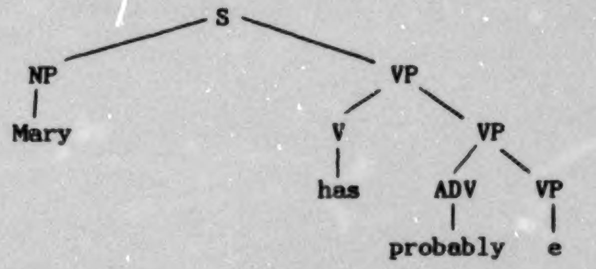
Jackendoff's (1972, 1977) analysis would encounter the same problem, because his analysis also claims that necessarily equal four is not a constituent. Thus, sentences such as 106 are problematic for the analyses of Jackendoff (1972, 1977) and Gazdar et al. (1982), but immediately accounted for by the S-Adverb basic rule, given the 'VP Deletion' metarule of Gazdar et al. (1982).

Given the VP Deletion analysis and the S-Adverb basic rule, it would seem that ungrammatical strings such as that in 111 would be generated.

111. *John has probably gone to Cleveland and Mary
has probably, too.
(with no pause between the auxiliary verb and adverb)

The second conjunct would apparently be assigned a structure as in 112.

112.



The VP Deletion metarule allows the rule $[\text{VP} \rightarrow \text{V} \ \underset{[+\text{null}]}{\text{VP}}]$. The S-Adverb basic rule allows $[\underset{[+\text{null}]}{\text{VP}} \rightarrow \text{ADV} \ \ \text{VP}]$. The Head Feature Convention, assuming as do Gazdar et al. (1982) that null is a head feature, ensures that the lower VP is also [+null]. We could account for the ungrammaticality of sentences such as 111 simply by revising the S-Adverb basic rule as in 113, specifying that the dominating VP must be [-null].

```
113.  VP -> ADV  VP
         [-null]
```

However, in section 6, I will argue that it is not necessary to specify that the dominating VP is [-null] in order to account for the ungrammaticality of sentences such as 111.

3.2 Evaluative and modal adverbs in clause-initial and clause-final positions

Evaluative and modal adverbs occur in clause-initial and clause final positions with comma intonation, as the following examples show.[6]

114. Unfortunately, John has been in an accident.
115. Obviously, he was driving while he was drunk.
116. He was not seriously injured, fortunately.
117. He will be released soon, probably.

They also occur in initial and final positions in embedded clauses.

118. Mike knows that unfortunately John has been in an accident.
119. The legend that Milton was an unpopular poet has lived so long that probably it will never be destroyed.[7]

The adverb must follow the complementizer or it will not be interpreted as part of the embedded clause. The only interpretation of 120 for example, is one in which <u>unfortunately</u> has scope over the root sentence and not just the embedded clause. Such sentences require parenthetical intonation.

120. Mike knows, unfortunately, that John has been in an accident.

Gazdar et al. (1982) mark <u>that</u>-clauses with the feature [+Complementizer]. <u>That</u>-clauses are introduced by the basic rule in 121.

```
121.  <6, [S -> that S], S'>
          [+C]      [-C]
```

The ID rule in 122 accounts for evaluative and modal adverbs in clause-initial and clause-final positions. Marking S as [-Complementizer], ensures that the adverb will not be part of the embedded clause if it precedes the complementizer.

```
122.  <2, [S -> ADV, S], ADV'( S')>
                    [-C]
      where ADV = the evaluative and modal adverbs
               [2]
```

We now have two lexical classes consisting of modal and evaluative adverbs. Adverbs of lexical class 1 are introduced by the ID rule <1, [VP -> ADV VP], ADV'(VP)> and adverbs of class 2 are introduced by the ID rule in 122. It is, of course, desirable to relate adverbs of lexical class 2, which are of type <<s,t>,<s,t>>, with their doublets in lexical class 1. The lexical rule in 123 will accomplish this (cf. Dowty (1978) for an explanation of lexical rules in Montague Grammar).

123. If a θ ADV [1], then $F_i(a) \in$ ADV [2], where $F_i(a)$ = a.
Translation: $\lambda {}^\wedge\mathcal{P} \lambda P[a'({}^\wedge\mathcal{P}(P))]$, where ${}^\wedge\mathcal{P}$ is a variable over VP type denotations and P is a variable over NP type denotation.

This rule states that if there is an adverb which is of lexical class 1, then there is a corresponding adverb of lexical class 2 which has the same form, and which is translated as $\lambda {}^\wedge\mathcal{P} \lambda P[a({}^\wedge\mathcal{P}(P))]$.

If 122 is revised so that the lexical class 2 includes other adverbs which occur clause-initially and finally, then there will not be a one-to-one correspondence between the two clases of adverbs. "Style-disjuncts" (cf. Greenbaum (1969) and Schreiber (1972)), such as confidentially, honestly, and frankly in the following examples, occur in clause-initial and clause-final positions, but do not occur in pre-verbal positions without parenthetical intonation.

124. Confidentially, she's no friend of mine.
125. Honestly, I didn't mean to insult you.
126. Frankly, I simply don't like you.

Temporal adverbs, such as yesterday and tomorrow, also occur in clause-initial and clause-final positions, but not pre-verbal positions. Some frequency adverbs, such as occasionally and frequently, occur in clause-initial and clause-final positions, and also occur in pre-verbal positions. It can be assumed that these adverbs are in lexical class 2, but, as explained in section 4, it will still be necessary to posit a distinct category to which these adverbs belong when in pre-verbal positions. A few VP adverbs, such as quickly and slowly, also occur in clause-initial and clause-final positions. All VP adverbs occur in pre-verbal positions, but generally do not occur in pre-verbal position when the verb is an auxiliary. Thus, if quickly or slowly occur in pre-auxiliary position, then it is reasonable to assume that they occur in this position by virtue of belonging to lexical class 1. Speakers differ as to whether or not they accept these adverbs in pre-auxiliary positions; some speakers accept sentences such as 127 and 128, while others do not.

127. The man quickly will bang the drum.
128. The children slowly have recited the alphabet.

This difference can be accounted for if we assume that for some speakers, these adverbs are members of lexical class 1, but for other speakers, they are not.

FOOTNOTES

1. I am using the type assignments for NP's and VP's given in Klein and Sag (1982). In Klein and Sag (1982), VP's are third order predicates. Dowty (1979) and Bach (1980) have argued that modal and tensed VP's should be analyzed as third order predicates. As noted by Dowty (1980), once tensed and modal VP's are defined as third order predicates, "an expression of the category PredP [i.e. a tensed or modal VP] has the meaning of the subject of the sentence within its scope. Hence an adverb like *possibly*, which is part of the PredP, can likewise have the subject within its scope, which is the crucial semantic property that S-adverbs must have" (p.7).
2. Pullum (1982) gives arguments for the claim that *to* is a verb.
3. "Style disjuncts" (cf. Greenbaum (1969) and Schreiber (1972)), such as *confidentially*, *honestly*, and *frankly* in the following examples, have been labeled sentential adverbs, but do not occur in preverbal positions without parenthetical intonation.

    Confidentially, I wouldn't trust him.
    Honestly, I love that color on you.
    Frankly, I can't stand divinity fudge.
4. It should be noted that the rule of minor right node raising, [VP -> VP/NP NP], is not produced by the rightward displacement schema of Gazdar (1981), repeated below, because it is required that a be a clausal category.

    RDS: <9, [a -> a/B B], h_B[(a/B)'](B')>
    where a ranges over clausal categories
    and B can be any phrasal or clausal category.
5. Note that a sentence such as *Ed said he would catch and kill the rabid dog and catch and kill the rabid dog, he will.* cannot be used to motivate the minor right node raising rule, given the rule [V -> V and V] which Gazdar (1982) assumes. Given this rule *catch and kill the rabid dog* can be generated as a VP consisting of a complex V and an NP. Rather than argue against this complex V analysis, I have given an example in which it is clear that two verbs have not been conjoined, since *try to kill* is not a verb.
6. We cannot account for adverbs in clause-initial position by allowing the Topicalization Schema to apply to adverbs. If the Topicalization Schema were allowed to apply to adverbs, it would incorrectly be predicted that i and ii have readings in which the adverb has scope over only the lower clause.

    i. Probably, Jake knows that Howard will leave.
    ii. Unfortunately, Jake knows that his brother has been in an accident.

    Since the Topicalization Schema is stated as not applying to lexical categories, topicalization of adverbs is correctly ruled out. It is also necessary to rule out topicalization of adverb phrases. Otherwise, we predict an interpretation of iii in which the ADVP has scope over the lower clause.

    iii. Quite possibly, Jake knows that Howard will leave.
7. This example is from Jacobson (1964), citing B. Ifor Evans, *A Short History of English Literature*.

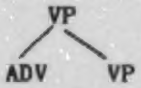
4. Temporal and Frequency Adverbs

In this section I will compare the placement of temporal and frequency adverbs with the placement of evaluative and modal adverbs. The results of this comparison will be important in the analysis of adverb stranding to be presented in section 6.

Adverbs which specify frequency, such as *always*, *sometimes*, *occasionally* and *usually*, occur in most of the same positions as evaluative and modal adverbs. Temporal adverbs, which specify a particular period of time, such as *yesterday*, *today*, and *tomorrow*, occur in much fewer positions.

Frequency adverbs, like evaluative and modal adverbs, occur in the configuration in 1 when immediately preceding a main or auxiliary verb.

1.



Examples such as 2 and 3 provide evidence for the higher VP node in 1.

2. Laurie said she would always love her mother, and
   always love her mother, she will. [VP Preposing]
3. Danny will always love Marsha, and Mark will, too.
   [VP Deletion]

If 2 and 3 are examples of VP Preposing and VP Deletion, as they certainly seem to be, then *always love her mother* and *always love Marsha* must be VP's. Examples such as 4 provide evidence for the lower VP node in 1. Because of the presence of *sometimes* in the second conjunct, the only reading of 4 is one in which only *paid for dinner* has been 'deleted' semantically, and not *usually paid for dinner*. If this is an example of VP Deletion, as it certainly seems to be, then *paid for dinner* in the left conjunct must be a VP.

4. Michael has usually *paid for dinner*, and Beth sometimes
   has.

Temporal adverbs such as *yesterday* and *tomorrow* do not occur in positions immediately preceding main or auxiliary verbs, as examples such as 5-7 show.

5. *John yesterday went to the beach.
6. *John will tomorrow go to the beach.
7. *John tomorrow will go to the beach.

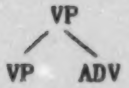
These adverbs, therefore, do not occur in the configuration in 1.

Temporal and frequency adverbs, unlike evaluative and modal adverbs, do occur at the right of conjoined VP's, as in 8 and 9.

8. Clark writes letters usually and sends telegrams sometimes.
9. Clark wrote a letter yesterday and sent a telegram today.

Such sentences can be accounted for if these adverbs occur in the configuration in 10.

10.



The following chart sums up the positions of occurrence of evaluative, modal, frequency, and temporal adverbs when not clause initial or clause-final. The phrase structure rules by which each type of adverb must be introduced are also listed.

	Left sister of VP	Right sister of VP	P-S rules
Evaluative and modal	yes	no	VP -> ADV VP
Frequency adverbs	yes	yes	VP -> ADV VP VP -> VP ADV
Temporal adverbs	no	yes	VP -> VP ADV

As shown, frequency adverbs occur as both left and right sisters of VP. Temporal adverbs occur only as right sisters of VP, while evaluative and modal adverbs occur only as left sisters of VP (cf. Section 3). In order to account for these observations, it is necessary to include the three ID rules and two LP rules given below.

11. <1, [VP -> VP, ADV] ADV'(VP')>
where ADV = the evaluative and modals
[1]
12. <2, [VP -> VP, ADV], ADV'(VP')>
where ADV = frequency adverbs
[2]
13. <3, [VP -> VP, ADV], ADV'(VP')>
where ADV = temporal adverbs
[3]
14. VP < ADV
[3]
15. ADV < VP
[1]

Given these ID and LP rules, the following phrase-structure rules will be basic rules of the grammar.

16. <1,[VP -> ADV VP], ADV'(VP')>
17. <2,[VP -> ADV VP], ADV'(VP')>
18. <2,[VP -> VP ADV], ADV'(VP')>
19. <3,[VP -> VP ADV], ADV'(VP')>

The ID rules in 11-13 cannot be collapsed in any way, because of the need to enforce different linear precedence restrictions on the three types of adverbs. I have used the rule numbers in the LP rules. Some other feature could have been used, but we would still need three distinct ID rules, since the adverbs in the ID rules would have to be marked with different features.

When Gazdar and Pullum (1981) introduced the ID/LP format, they did not specify any requirements on the form of LP rules. In Gazdar and Pullum (1982:21), it is claimed that the LP rule in 20 is a rule of English.

20. [+LEXICAL] < [-LEXICAL]

Thus, it is clear that Gazdar and Pullum intend features other than syntactic category features to be permitted in LP rules. Since "the rule number is assigned, by convention, to be the value of the feature LEXICAL" (Gazdar and Pullum (1982:17)), the use of rule numbers in LP rules is certainly not ruled out by the theory as stated. It should be noted that the rules in 18 and 19 are inconsistent with Gazdar and Pullum's generalization in 20. However, if we replace ADV with ADVP (adverb phrase) in these rules, they are no longer inconsistent with 20. I will leave open whether such a move should be taken, but note that if we replace ADV with ADVP in 16-19, we can account for the difference in positions of occurrence of these three kinds of adverbs by distinguishing three categories of ADVP, as well as three lexical classes of ADV's.

5. VP Adverbs

In this section an analysis of the placement of VP adverbs is presented. It is argued that VP adverbs occur as sisters to V, rather than, as Jackendoff (1972, 1977) claimed, as aunts of V.

I will limit this discussion to those VP adverbs with the fewest restrictions on their occurrence in positions other than clause-initial and clause-final. This group includes *quickly*, *slowly*, *intensely*, *incessantly*, *thoroughly*, *seriously*, *firmly*, *diligently*, *completely*, *tremendously*, *purposefully*, as well as *willingly*, *knowingly*, and *cleverly*.[1]

It is uncontroversial that VP adverbs are dominated by VP. It must be argued, however, that they occur as sisters, rather than aunts of V. Examples of 'VP Deletion' provide evidence for this claim.

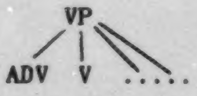
1. John has been seriously wounded, and Mary has been, too.
2. George has quickly read the book, and Mary has, too.
3. George had firmly refused the offer, and Mary had, too.

The only interpretation which these sentences have is one in which the adverb is included as part of what has been 'deleted' semantically. The only interpretation of 1, for example, is one in which *seriously wounded* has been 'deleted' semantically. In section 3.25 it was shown that a 'deleted' VP in a right conjunct could correspond to a VP in the left conjunct dominating an S-adverb and a VP or it could correspond to the VP which is a sister of the S-adverb. In 4, the most natural reading is one in which the VP dominating the ADV and following VP has been deleted. In 5, the only reading is one in which the VP which is a sister of the adverb has been deleted.

4. Two plus two will necessarily equal four and one plus one will, too.
5. John will probably go to Baltimore, and Mary definitely will.

If the adverbs in 1-3 were aunts of the V, we would expect these sentences to be ambiguous between an interpretation in which the higher VP has been 'deleted' semantically and one in which only the lower VP has been 'deleted'. Sentence 1 should be ambiguous between an interpretation in which *seriously wounded* is 'deleted' and one in which *wounded* is 'deleted'. However, this is not the case. This observation can be accounted for if there is no lower VP to which the adverb is a sister. In the following discussion, I will assume that VP adverbs are to be sisters of V, as in configuration 6.

6.



VP adverbs occur before main verbs, but not before auxiliary verbs, as the following examples show.

7. George quickly read the book.
8. *George quickly has read the book.
9. George has quickly read the book.

10. *George will quickly have read the book.
11. George will have quickly read the book.

Jackendoff (1972:75) marks examples such as 10 with a question mark, rather than an asterisk. However, my informants consistently rejected sentences such as 10. Some speakers accepted sentences such as 12 with *quickly* preceding the auxiliary verb, but only with an interpretation in which John was quick to bang the drums, not with the interpretation that the banging was quick.

12. John quickly has banged the drums.

VP adverbs also occur before prepositional phrases within the VP, as in 13, and in VP-final position, as in 14.

13. John gave the book quickly to Mary.
14. John gave the book to Mary quickly.

They do not occur before noun phrases within the VP, as 15-17 show, or before VP or S complements of the verb, as 18-20 show.

15. *John gave quickly the book to Mary.
16. *John gave quickly Mary the book.
17. *John persuaded quickly Kim to leave.
18. *John persuaded Kim quickly to leave.
19. *John wanted Kim quickly to leave.
20. *John promised Kim quickly that he would visit her.

In order to account for the data presented above, I will adopt the metarule in 21, as well as three LP rules.[2]

```
21.  <8, [VP -> V,      X], V'(F)>  ==>
               [-AUX]                               where ADV
     <8, [VP -> ADV, V,      X], ADV'(V'(F))>       = the VP adverbs
                    [-AUX]
```

This metarule states that for every ID rule of the grammar which expands VP as V and any categories X, there is also an ID rule which expands VP in exactly the same manner except ADV is also a daughter of VP. This metarule captures the generalization that VP adverbs may occur as daughters of any VP. I have given a semantic translation in which the semantic value of the adverb is a function from verb phrase type denotations to verb phrase type denotatins. The ID rules which will be the output of the metarule in 21 will have two lexical categories, ADV and V, as daughters of VP. Rule numbers, by convention, are features on the lexical category introduced by an ID rule, but in these rules two lexical categories are introduced. In order to ensure that the ID rules which are output by 21 will have V, and not ADV, marked with the rule number, I will adopt a convention that only the lexical category which is marked with the rule number in the ID rules which are output by 21 will have V, and not ADV, marked with the rule number. I will also assume that lexical categories which are not marked with a rule number by this convention may have a rule number specified as one of their features; otherwise it would not be possible to specify which adverbs occur in the ID rules output by metarule 21. Metarule 21 will be revised as in 22.

22. Revised VP metarule
<8, [VP -> V, X], V'(F)> ==>
<8, [VP -> ADV$_{[6]}$, V, X], (ADV'(V'))(F)>
where ADV$_{[6]}$ = the set of VP adverbs

In order to rule out ungrammatical sentences such as 18-20, it is necessary to adopt the LP rules in 23 and 24.

23. VP < ADV$_{[6]}$
24. S < ADV$_{[6]}$

In order to rule out ungrammatical sentences such as 15-17, but allow grammatical sentences such as 7, it will be necessary to adopt a new type of LP rule. Note that 15-17 cannot be ruled out by adopting the LP rule in 25, since this rule would incorrectly rule out sentences such as 7, in which the adverb precedes the NP.

25. NP < ADV$_{[6]}$

What is needed is an LP rule which allows adverbs of class 6 to precede NP's, as in 7, but not to immediately precede NP's. LP rules, as originally conceived by Gazdar and Pullum (1981), cannot make a distinction between precedence and immediate precedence. By dis allowing such distinctions to be expressed by LP rules, the predictive power of the LP/ID format is enhanced. However, if the metarule analysis of VP adverbs is correct, it will be necessary to allow LP rules to make such a distinction. The rule in 26, which states that adverbs of class 6 may not immediately precede NP's will be needed.[3]

21. ADV$_{[6]}$ $I\not<$ NP

Footnotes

1. _Willingly_, _knowingly_, and _cleverly_, and other "passive-sensitive" adverbs, also belong to the same lexical class as evaluative and modal adverbs, since they occur in te same positions as these adverbs, as well as in the same positions as VP adverbs. _Willingly_, _knowingly_, and _cleverly_ are known as "passive-sensitive" adverbs, because of the difference in interpretation of examples such as i and ii.
   i. The doctor willingly examined Mary.
   ii. Mary was willingly examined by the doctor.

The passive sentence ii has a reading which the active sentence does not: ii has a reading in which Mary was willing, as well as a reading in which the doctor was willing, but i only has the reading in which the doctor was willing. Given that these adverbs belong to the same lexical class as evaluatives and modals, as well as VP adverbs, it will be possible to adopt Dowty's (1980) analysis of the semantics of passive sensitive adverbs (cf. footnote 2 below).

2. We can follow Dowty's (1980) treatment of passive-sensitive adverbs if we incorporate transitive verb phrases (TVP's) into the grammar (cf. Gazdar and Sag (1981) for a discussion of TVP's in GPSG) and adopt the metarule below in addition to metarule 21. (These two metarules could be collapsed into a single metarule.)

   <10, [TVP -> V, X], V'(F)> ==>
   <10, [TVP -> ADV, V, X], ADV'(V'(F))>
   [9]

   In Dowty's analysis passive-sensitive adverbs belong to the class PredP/PredP (which corresponds semantically to our lexical class 2), IV/IV (which corresponds to our lexical class 6), and TV/TV (which corresponds to our lexical class 9).

3. It has been pointed out to me by David Dowty that this is not the only possibility. The same data could also be accounted for by permitting rules consisting of disjunctions of LP rules. The rule below would yield the same results as rule 26.

   ADV < V or ADV > NP VP, S
   [6] [6]

6. Adverb Stranding

This section deals with sentences in which an adverb immediately precedes a 'deletion' or 'movement' site, and is thus stranded. It will be shown that given the rules for adverb placement proposed in previous chapters, and the Trace Introduction Metarule of Sag (1982), data involving adverb stranding is immediately accounted for.

Quantifiers, like sentential adverbs, cannot immediately precede 'deletion' or 'movement' sites. The analysis to be presented, unlike previous analyses, will account for the identical 'behavior' of sentential adverbs and quantifiers before movement and deletion sites.

In section 6.1 previous analyses of adverb (and quantifier) stranding will be discussed. In section 6.2 a rule will be proposed for quantifier placement and the Trace Introduction Metarule will be discussed. In section 6.3, it will be shown that the Trace Introduction Metarule interacts with the rules for adverb and quantifer placement to yield correct predictions concerning sentences in which adverbs and quantifiers immediately precede VP 'movement' sites. In this section, it will also be shown that, given certain assumptions about the feature _null_, correct predictions result concerning adverbs and quantifiers before VP 'deletion' sites.

6.1 Previous analyses

Baker (1981) discusses the ungrammaticality of sentences such as 1 and 2 in which a sentential adverb or quantifier precedes a deletion site.

1. *Fred has never been rude to Grandfather,
   but John has always ø.
2. *I have read Moby Dick, and they have all ø, too.

Baker assumes that adverbs and quantifiers precede the finite auxiliary verb in underlying structure. A transformational rule of Auxiliary Shift moves unstressed auxiliaries to the left of adverbs or quantifiers. The sentence in 3, for example, derives from 4 by Auxiliary Shift. Auxiliary Shift moves the stressless auxiliary _have_ to the left of the adverbs _probably_ and _never_.

3. George and Martha have probably never seen a real politician.
4. George and Martha probably never have seen a real politician.

According to Baker's analysis, the underlying structures for 1 and 2 would have to be 5 and 6, respectively.

5. Fred has never been rude to Grandfather,
   but John always has ø.
6. I have read Moby Dick, and they all have ø, too.

Baker claims that auxiliaries before deletion sites are always stressed. Since the auxiliary is stressed, Auxiliary Shift will not apply to 5 or 6, and thus 1 and 2 are predicted to be ungrammatical.

Baker's analysis hinges on the claim that only unstressed auxiliaries occur before adverbs or quantifiers. Sag (1980) and Ernst (1983) have cited the following counterexamples to this claim.

7. They denied that John has always admired Susan,
but he HAS always admired her. (Sag's example 6)
8. a. Do you mean to say in front of this committee,
sir, that every single factor has been taken into
account in your budget estimates?
b. Well...we HAVE probably glossed over the effects
of the FOOD PRICE increases. (Ernst 1983)

In 7 the auxiliary preceding the adverb in the second clause is stressed and the VP of the first clause is 'echoed' to some extent. In 8.b. the adverb follows a stressed auxiliary and a second stress occurs in the VP. Ernst (1983) discusses various discourse conditions under which sentences such as 8.b. are acceptable. He suggests that "adverbs may follow auxiliaries whenever discourse conditions allow the auxiliaries to be stressed--whether or not there is a second stress in the VP" (p.547). Thus, according to Ernst, the acceptability of sentences such as 7, as well as sentences such as 8.b., is dependent on the discourse situation. He concludes that "we should allow the grammar to generate adverbs freely after stressed auxiliaries, in addition to the regular cases of nonstressed auxiliaries. The only requirements are discourse conditions relating to appropriate structures for contrasts and to different degrees of stress" (p.547).

It is clear that Baker's analysis is incorrect and that the relation between auxiliary stress and adverb placement is governed by discourse conditions. It is not clear how to rule out sentences such as 1 and 2, which are ungrammatical whether or not the auxiliary is stressed.

Sag (1978, 1980) considers the ungrammaticality of such sentences to be related to the ungrammaticality of the sentences below. According to Sag, these sentences are ruled out by the generalization that "adverbs and so-called 'floated' quantifiers may not appear in surface structure in a position immediately preceding an extraction site" (1980:255).

9. *I don't know what they are all ∅.
*I don't know how happy they are ever ∅.
[WH movement]
10. *I know a first grader who has finished more lesson
units than the second graders have all ∅.
*The activists are now more active than they were ever ∅.
[Comparative Deletion]
11. *My brother has studied karate, and my sisters have all ∅,
also.
*I don't know if Leslie has ever studied karate, and I
don't know if Gwendolyn has ever ∅, either.
[VP Deletion]
12. *Sandy is polite to strangers, which I doubt very much
that your brothers are all ∅.
*Sandy is polite to strangers, which I doubt very much
that Ralph is ever ∅.
[Relativization]

13. *None of them were Communists, but Socialists, they were
all ∅.
*They used to be Socialists, but Communists, they were
never ∅.
[Topicalization]
14. *The more unhappy you say they are, the happier they
are all ∅.
*The more polite you tell them to be, the more polite they
are usually ∅.
[The-More-the-Merrier-Fronting]
15. *They said our children would be polite and polite they
are all ∅.
*They said our children would be polite, but polite, they
are never ∅.
[VP Preposing]

The surface filter in 16 is posited to account for the ungrammaticality of 9-15.

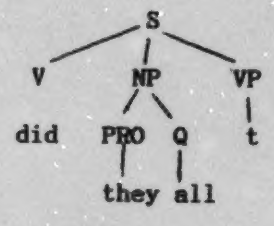
16. $*\left\{\begin{matrix}\text{Q}\\\text{ADV}\end{matrix}\right\}$ extraction site

Sag (1978) notes that this surface filter must be modified in view of the grammaticality of questions such as 17 and 18.

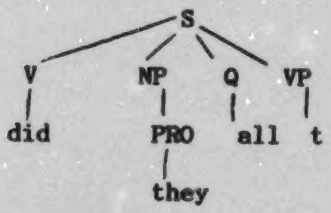
17. Did they all ∅?
18. Does he usually ∅?

He sketches a solution to the problem presented by questions such as 17. Such questions, he claims (following Postal (1974) and Maling (1976)) have two constituent structures: one in which PRO and Q form a constituent (NP), as in 19, and another in which they do not, as in 20.

19.



20.



In 19 Q is not the sister of an extraction site, but in 20 it is. Sag accounts for the grammaticality of questions such as 17 by revising the filter as in 21.

21. * $\left\{\begin{matrix} \text{Q} \\ \text{ADV} \end{matrix}\right\}$ extraction site

where Q, ADV is a sister of the extraction site

This filter rules out an adverb or quantifier before an extraction site only if it is a sister of the extraction site. Sag obviously intends "sister of an extraction site" to be taken to mean sister to a node which immediately dominates a trace or null element. This filter will not rule out the grammatical sentence in 17, because 17 has a structure, 19, in which the Q is not a sister to the extraction site. Note that sentences such as 22 are correctly ruled out by the filter in 21 because they have only a constituent structure in which the Q is a sister of the extraction site.

22. *Did the men all?

Sentences such as 18 will still be incorrectly ruled out, however, since there is no evidence for a constituent structure in which the adverb is not a sister of the extraction site. Sag (1978) claimed that examples such as 18 were grammatical only if a pronoun precedes the adverb, and cited the following examples which contrast with 18.

23. *Does President Carter usually ∅?
24. *Will Anita Bryant ever ∅?

However, my informants judged examples such as 25 to be perfectly acceptable. Perhaps the unacceptability of 23 and 24 has to do with the use of proper names.

25. Do your friends usually?
(with no pause before *usually*)

Examples such as 26, from Baker (1981), will also be incorrectly ruled out by Sag's filter, as well as 27 and 28 from Ernst (1983).

26. He's gotten along well with Fred in the past few weeks,
but he hasn't always.
27. Terry knows how to build an H-bomb.
No--does he REALLY??
28. Joe says he will run a four-minute mile on a steeple-chase course.
How could he POSSIBLY?!!

Ernst (1983) notes that the counterexamples to Sag's surface filter involve a restricted set of adverbs: time adverbs, such as *usually*, *sometimes*, *then*, *now*, *recently*, *soon*, and the two adverbs *really* and *possibly* (for some speakers). However, there is another type of counterexample in which VP adverbs, such as *quietly*, *partially*, and *slowly*, apparently are sisters to deletion sites. These examples involve the verbal ellipsis phenomenon known as "gapping".

29. John will loudly answer my questions and
    Mary ø quietly ø.
30. Todd has thoroughly read the book and
    Mark ø partially ø.

In examples 29 and 30, the auxiliary, the verb, and object NP are 'missing'. Whether the VP adverbs are sisters to a VP which dominates the verb and object or are within the VP and sisters to the V, they are sisters of extraction sites. If the adverb is a sister to VP, one might consider arguing that 29 and 30 are actually examples of VP Deletion—that, for example, *will answer my questions* has been 'deleted' before *quietly* and nothing has been 'deleted' after the adverb. However, evidence can be given to the contrary. VP Deletion can apply within an embedded S, as in 31, but gapping cannot; although 32 is grammatical, the sentences in 33 (from Sag 1977) are not.

31. John will go to the movies and I know that Bill will, too.
32. Alan went to New York and Betsy ø to Boston.
33. *Alan went to New York, and
    a. I know (that)
    b. it seems (that)
    c. Bill met a man who claimed (that)
    Betsy ø to Boston.

If 29 and 30 were examples of VP Deletion, we would expect sentences such as 34 and 35 to also be grammatical, but they are not.

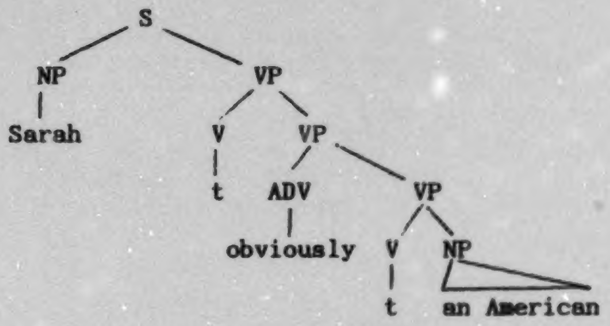
34. *John will loudly answer my questions and I know that
    Mary quietly.
35. *Todd has thoroughly read the book and I know that
    Mark partially.

The grammaticality of sentences such as 29 and 30 can be accounted for if it is assumed that the surface filter applies only to sentential adverbs. (Ernst seems to assume that this is what Sag intended anyway.) There are similar examples of Gapping involving sentential adverbs, but in these cases the adverbs are not sisters of the extraction sites.

36. Olga will probably marry a Russian and Sarah ø obviously
    ø an American.

Given the analysis of modal adverbs in chapter 3, the right conjunct in 36 will be assigned the structure in 37.

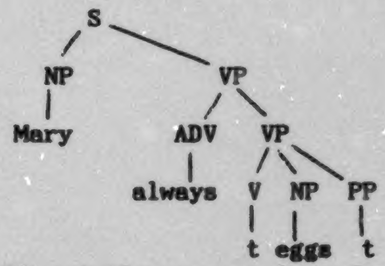
37.



The sentence in 36 is a counterexample to the original filter in 16. It is not a counterexample to the revised filter in 21, however, since obviously is not a sister of the extraction site.[1] There are also examples of frequency adverbs preceding 'gapped' verbs, as in 38. I have argued that such adverbs are sisters of VP. If this is the case, then the adverb in 38 is not the sister of an extraction site, but but instead the aunt, as in 39.

38. John usually eats cereal for breakfast and
    Mary always ø eggs ø.

39.



Given that 36 and 38 are examples of gapping, they count as evidence against the original filter in 16, but not the revised filter in 21.

To sum up the discussion of filters: the following counterexamples to the original filter have been given:

40. Do they all?
41. Olga will probably marry a Russian and Sarah ø
    obviously ø an American.
42. John usually eats cereal for breakfast and
    Mary always ø eggs ø.
43. ...does he REALLY?
44. ...How could he POSSIBLY?!!
45. Does he usually?
46. He's gotten along well with Fred in the past few weeks,
    but he hasn't always.
47. John will loudly answer my questions and Mary
    ø quietly ø.
48. Todd has thoroughly read the book and Mark ø
    partially ø.

Once the surface filter is revised, as in 21, the examples in 40-42 will no longer be counterexamples. However, the examples in 43-48 are apparent counterexamples to the revised filter. I will have nothing to say about 43 and 44. The other examples involve either time adverbs or

VP adverbs occuring before deletion sites. The correct observation seems to be that sentential adverbs (excluding time adverbs) cannot occur as sisters of extraction sites. While a surface filter can be devised to account for this observation, I would claim that such an analysis is misguided. Not only does it fail to explain why Q's and S-ADV's should 'behave' alike (i.e. why both should fail to occur as sisters of extraction sites), but, within the framework to be presented, it is unnecessary.

Baker (1981) and Ernst (1983) have both claimed that Sag's surface filter analysis "seems rather implausible from the point of view of language acquisition" (Ernst, p. 547). Ernst proposes that Sag's filter be replaced by a filter which forbids material between auxiliary verbs and a VP-deletion site. However, as Ernst points out,such a filter incorrectly rules out examples of Subject-Auxiliary Inversion such as that in 49.

49. Phil was diving into a wet dishrag.
WAS he ø?! (Ernst 1983)

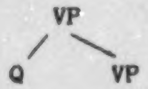
In the following sections, an analysis will be offered within a version of Generalized Phrase Structure Grammar. In this analysis, no special constraint is needed to rule out ungrammatical sentences in which a sentential adverb (excluding time adverbs) or a quantifier is a sister of an extraction site (actually what is to be excluded is a structure in which the node S-ADV or the node Q is a sister of an extraction site). Certain independently motivated aspects of the grammar interact to produce the desired results. Since no surface filter will be required, this analysis is compatible with Jacobson's (1982) tentative claim that "no constraint in the grammar can explicitly mention gaps" (p. 207). Under a filter analysis it just happens to be the case that both Q's and S-ADV's cannot be sisters of extraction sites. Under the analysis to be presented, Q's and S-ADV's 'behave' alike in this respect because of a structural identity.

6.2 Assumptions Underlying the Structural Analysis

In the structural analysis to be presented, I will assume the rules for introducing S-ADV, frequency, temporal, and VP adverbs given in the preceding chapters.

I will also assume that 'floated' quantifiers occur in the configuration in 50. Baltin (1982) argues for such a structure.

50.



Examples of VP Preposing and VP Deletion provide evidence for the higher VP node in 50.

51. They said that they would all work on that, and all work on that, they did. (Baltin 1982, example 36)
52. They said they will all work on that and they will.

Examples such as 53 provide evidence for the lower VP node in 50.

53. The women will all go to Rapid City and Howard will, too.

Quantifiers in pre-verbal positions will be introduced by the basic rule in 54.

54. <10, [VP -> Q VP]>
where Q(10) = all, each, both

I have not given a semantic rule in 75. It is necessary to provide a semantic analysis of sentences with 'floated' quantifiers which is compatible with the syntactic rule in 54, if this analysis is to be viable. However, since the evidence for the syntactic rule in 54 is compelling, I will assume for now that such a corresponding semantic rule can be motivated.

In the analysis to be presented it will also be assumed that all traces are introduced by means of the Trace Introduction Metarule (TIM) of Sag (1982). The TIM does much the same work for GPSG which the Immediate Dominance Principle of Sag (1977) did for TG. The TIM replaces linking rules such as 55.

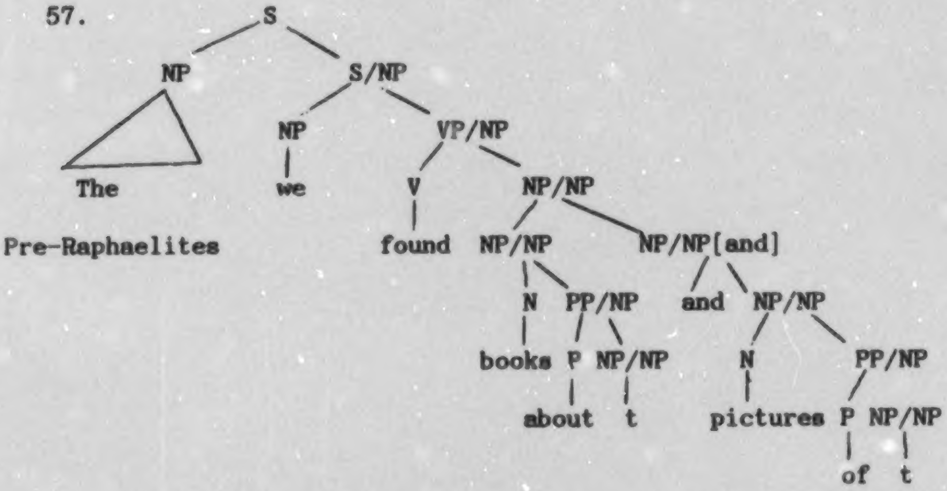
55. [NP/NP -> t]

The TIM was proposed in order to avoid problems for Gazdar's (1981) treatment of coordination which were due to the use of linking rules such as 55. Gazdar's coordination schema in 56 allows for coordination of NP/NP's as in 57.

56. $\langle 2, [a \rightarrow a_1 \ldots \underset{[B]}{a_n}], B'(a_1', \ldots, a_n') \rangle$
where B $\in$ [and, or] and a is any syntactic category

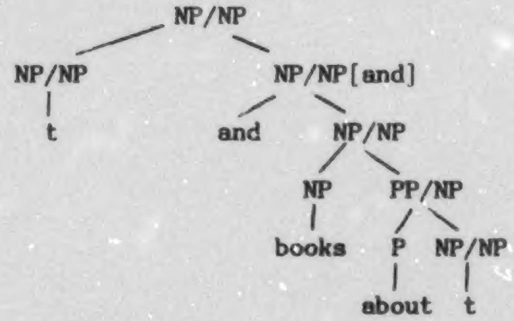
$\langle 3, [a \rightarrow B\ a], a' \rangle$
where B $\in$ [and, or,...] and a is any syntactic category

57.

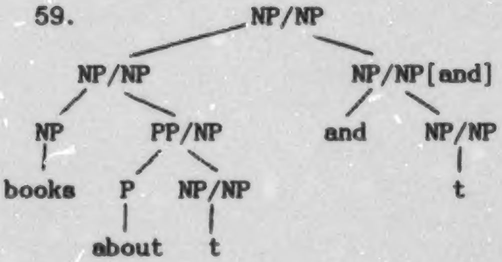


Given the linking rule in 55 and Gazdar's coordination schema, subtrees as in 58, 59, and 60 (Sag's 13 a,b,c) will be allowed and the ungrammatical sentences in 61, 62, and 63 (Sag's 14 a,b,c) will be generated.

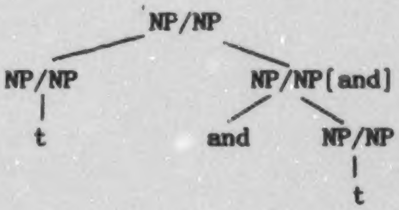
58.


NP/NP
NP/NP
t
NP/NP[and]
and
NP/NP
NP
books
PP/NP
P
about
NP/NP
t

59.


NP/NP
NP/NP
NP
books
PP/NP
P
about
NP/NP
t
NP/NP[and]
and
NP/NP
t

60.


NP/NP
NP/NP
t
NP/NP[and]
and
NP/NP
t

61. *The Pre-Raphaelites, we found [[t] and [books about t]].
62. *The Pre-Raphaelites, we found [[books about t] and [t]].
63. *The Pre-Raphaelites, we found [[t] and [t]].

In each of the subtrees in 58-60 a trace has been introduced (by the linking rule in 55) under a slash category node which is identical to the node immediately dominating it. What is needed is a means of introducing traces which will not allow them to appear under a node which is identical to the node immediately dominating it. The TIM in 64 accomplishes this by the condition that a not equal B.

64. TIM:

[a/B ->...B/B...] ==> [a/B ->...t...]
where a ≠ B

If 55 is replaced by TIM, the ungrammatical sentences in 61-63 will no longer be generated. Sag (1982:333) states that 61 will not be generated because "TIM would have to produce rules like the one in (17) [65 below]."

65. [NP/NP -> t NP/NP]
[and]

"This could only happen if the coordination schema ... were taken as input to TIM. But on independent grounds (see Gazdar (in press)), metarules may not operate on nonfinite schemata" (p.313). But note that even if the TIM were allowed to operate on the coordination schema, rule 65 would not be produced by TIM. Rule 65 could only result if the TIM took 66 as its input, but 66 is not a possible input to TIM since the condition that a does not equal B is not met.

66. [NP/NP -> NP/NP NP/NP]
[and]

Sentences 62 and 63 are ruled out by the a ≠ B condition. As Sag states, to generate 62, "TIM would have to apply so as to produce the rule in (18) [67 below]" (p.333).

67. [NP/NP -> and t]
[and]

"However this is impossible, as the input rule here would be the rule in (19) [68 below] ... which violates the a ≠ B condition on TIM" (p.334)

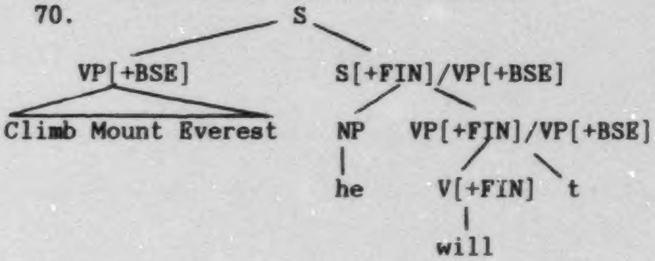
68. [NP/NP -> and NP/NP]
[and]

To generate 63, "one would need both rules (17) [67 above] and (18) [68 above] (p.334), which are, of course, ruled out."

Sag does not specify exactly what is meant by the condition a ≠ B. In the cases Sag discusses a ≠ B could be taken simply to mean that a is not the same category as B. However, such a condition will lead to incorrect predictions concerning VP Fronting. Given the analysis of VP Fronting presented in Gazdar et al. (1982), the tree for the sentence in 69 will be 70.

69. Climb Mount Everest, he will.

70.



The node dominating the trace in all cases of VP Fronting would be VP/VP. If the a ≠ B condition is taken to mean only that a and B may not have the same syntactic category features, then VP Fronting would be ruled out.

The TIM can be reconciled with VP Fronting, if we assume that a equals B only if the two nodes a and B are identical with respect to all features, the major category class simply being one of these features. Thus, a does not equal B if one or more features differ. Given this interpretation of TIM, VP Fronting will be allowed since the two VP's of the VP/VP node will differ in their features.

6.3 VP 'Deletion' and 'Movement'

In this section I will deal with cases in which an adverb or quantifier precedes a VP extraction site. These cases include examples of Comparative Deletion, VP Topicalization (VP Fronting), The-More-the Merrier Fronting, and VP Deletion. I am assuming that adjective phrases following the copular be are VP's marked as [+PRED], as do Gazdar et al. (1982). I will restrict the discussion to VP Preposing and VP Deletion, since it is not clear how examples of Comparative Deletion or The-More-the-Merrier Fronting should be handled.

6.31 VP Fronting

In Gazdar et al. (1982) VP Fronting is accounted for by the Topicalization schema below in conjunction with the slashing mechanism presented in Gazdar (1981) to account for unbounded dependencies.

71. <13, [S -> a S/a], λh_a [(S/a)'] (a')
when a = VP, then a is to be [-FIN, -INF, -ASP]

What is relevant to our discussion is the syntactic rule in the schema in 71 (i.e. [S -> a S/a]). The syntactic rule states that an S may consist of a phrasal category a followed by an S which is 'missing' an a. The slashing mechanism ensures that the VP which is 'missing' from S has the same features as the VP which is topicalized. Given the assumption that adverbs and quantifiers are 'Chomsky-adjoined' to VP's and that traces are introduced by TIM, I will show that the ungrammaticality of sentences such as 72 and 73 is predicted.

72. *John said he would pay me and pay me he will definitely ∅.
(with no pause between will and definitely)

73. *They said they would all pay me and pay me they will all ∅.

As in chapter 3, S-Adverbs in pre-verbal positions will be introduced by the basic rule repeated in 74. The slashing mechanism applies to this basic rule to give the derived rule in 75.

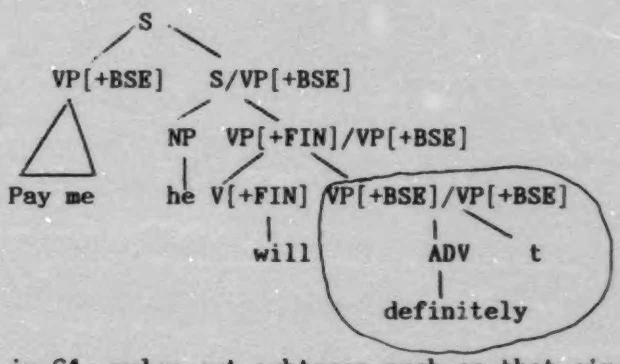
74. <1, [VP -> ADV VP]>
75. <1, [VP/VP -> ADV VP/VP]>

As stated earlier, 'floated' quantifiers will be introduced by the rule in 76. The slashing mechanism will apply to yield the rule in 77.

76. <10, [VP -> Q VP]>
77. <10, [VP/VP -> Q VP/VP]>

The tree for 72 would have to be 78. (The distribution of features is explained below.) The sentence in 73 would have the same structure.

78.



But the TIM, given in 64, rules out subtrees such as that circled in 78, since traces are only allowed as daughters of nodes a/B where a does not equal B. In the circled subtree the node immediately dominating the trace is an a/B where a is identical to B (i.e. a = B). Given the TIM, the tree in 78 is not a possible structure.

It can be shown that in any instantiation of the rule [VP/VP -> ADV VP/VP] in which the VP's of the dominated VP/VP node are identical, the VP's will all share the same features, and thus a will always equal B. To see why this is so, consider the subscripted version of the derived rule below.

79. [VP_1/VP_2 -> ADV VP_3/VP_4]

The Head Feature Convention will ensure that VP_1 and VP_3 have identical features, since VP_3 is the head of VP_1. The TIM will require that VP_3 and VP_4 have the same features if a trace is to be introduced at this node (they would be the B/B in the TIM). Since VP_1 has the same features as VP_3 which has the same features as VP_4, VP_1 must have the same features as VP_4. The slashing mechanism will require that VP_2 and VP_4 have identical features. Thus VP_2 and VP_1 both have the same features as VP_4 and are therefore identical.

Given the TIM, then, it is impossible for a trace to appear as the sister of an ADV or Q which is introduced by one of the basic rules in 74 or 75 or one of the derived rules in 75 or 76. It is, therefore, predicted that S-Adverbs and quantifiers will not immediately precede the site of a 'moved' VP.

In section 5 I argued that VP adverbs are introduced by the metarule repeated in 80. Given this assumption, it is obvious why VP adverbs cannot immediately precede a 'moved' VP--they do not have VP's as sisters.

80. <n, [VP -> V, X], V'(F)> ==>
<n, [VP -> ADV, V, X], ADV'[V'(F')]>

Frequency adverbs apparently occur before 'moved' VP's as in 81.

81. John said he would always love his mother and love his mother, he will always.
(without a pause between will and always)

In section 4 I argued that frequency adverbs are introduced by the basic rules in 82 and 83.

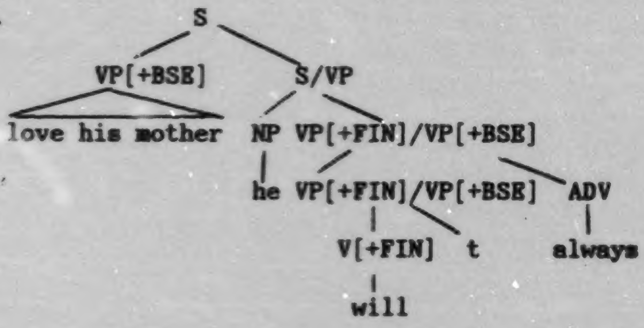
82. <3, [VP -> VP ADV]>
83. <3, [VP -> ADV VP]>

The slashing mechanism applies to 82 to yield the derived rule in 84.

84. <3, [VP/VP -> VP/VP ADV]>

Given the derived rule in 84, the grammaticality of sentences such as 81 is predicted. The second conjunct in 81 will be assigned the structure in 85.

85.



The slashing mechanism applies to the basic rule in 86 (cf. Gazdar et al. (1982)) to yield the derived rule in 87.

86. [VP -> V VP]
[+FIN] [+BSE]
87. [VP / VP -> V VP / VP]
[+FIN][+BSE] [+BSE][+BSE]

The rule in 87 can serve as input to the TIM since a does not equal B (i.e. VP[+FIN] is not identical to VP[+BSE]) and the rule in 88 will result.

88. [VP/VP -> V t]
[+FIN][+BSE]

Thus, the rules in 84 and 88, the slashing mechanism, and the Topicalization schema will interact to predict the grammaticality of sentences, such as 81, in which a frequency adverb immediately precedes a VP 'movement' site.

Since temporal adverbs are also introduced as right sisters of VP, we would expect sentences such as 89 to be grammatical, but they are not.

89. *John said he would go to the store tomorrow, and
go to the store, he will tomorrow.
(with no pause before the adverb)

It is not clear to me how to explain the ungrammaticality of such sentences. If we are to maintain that temporal adverbs are introduced by the syntactic rule [VP -> VP ADV], as argued in chapter 4, it will be necessary to provide an account of this data. If temporal adverbs were instead introduced by the rule [S -> S ADV], as evaluative and modal adverbs are, the ungrammaticality of 89 would be accounted for. However, it then becomes difficult to explain why these adverbs are not preceded by a pause as evaluative and modal adverbs are.

To summarize this section, the basic rules and TIM, along with other motivated rules of the grammar, interact to give correct predictions concerning quantifiers, S-Adverbs, frequency adverbs, and VP adverbs before VP 'movement' sites. It is not yet clear how to account for predictions involving temporal adverbs before VP 'movement' sites.

6.32 VP Deletion

S-Adverbs and VP adverbs do not occur before 'deleted' VP's, as the following examples illustrate.

90. *John has probably gone to Cleveland and
    Mark has probably ø, too.
    (with no pause between the auxiliary and adverb)
91. *Karen has thoroughly read this book and Doris
    has thoroughly ø, too.

Quantifiers do not occur before 'deleted' VP's, unless immediately preceded by a pronoun.

92. *The men have all left for lunch and the women
    have all ø, too.
93. *Have the men all ø?
94. Have they all ø?

Frequency adverbs apparently appear before 'deleted' VP's, whether or not a pronoun precedes. (I will argue below that the adverb is actually following the deleted VP, not preceding it.)

95. John has been nice to me lately, but he hasn't always.
96. Does he usually?
97. Do your friends usually?

As claimed in chapter 4, temporal adverbs do not precede VP's, and thus do not precede VP 'deletion' sites.

Gazdar et al. (1982) give the following metarules to account for 'VP Deletion':

```
98.  VPD:   <[VP -> V  VP], F >    ==>
            [+AUX]
            [-PRP]
            [-GER]

            <[VP -> V  VP], F >
                       [+NUL]
```

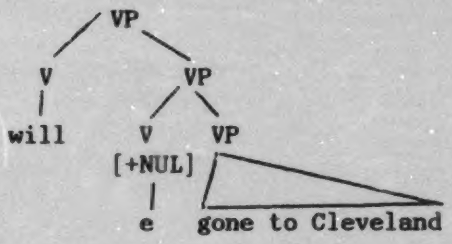
The metarule in 98 "takes any V [+AUX, -PRP, -GER] rule which expands as V followed by V', and simply adds the feature +NULL to the complement V'" (Gazdar et al. p. 606). The rule in 99 introduces <u>e</u>, which represents the empty string.

99. <16, [VP -> e], v>
[+NUL]
where v is a contextually bound variable ranging over VP denotations.

It is important to determine what kind of feature null is. Gazdar and Pullum (1982) discuss feature instantiation principles and distinguish two types of features—head features and foot features. Null is obviously not a head feature. If it were, the V of a [+null] VP would also be [+null] (since V is the head of VP and, by the Head Feature Convention, heads must have head features identical to those of their mother node), and ungrammatical sentences such as 100 would be produced. The right conjunct in 100 will be assigned the structure in 101.

100. John will have gone to Baltimore and
Betty will e gone to Cleveland.

101.



Since null is not a head feature, it may be a foot feature. Gazdar and Pullum (1982:34) note that "there are foot features that are explicitly specified in listed ID rules, or which have arisen through the operation of metarules. Such foot features are inviolate and cannot be copied or otherwise tampered with in the feature instantiation mapping." What is important for our purposes is the claim that foot features which are explicitly specified in the syntactic rules are not subject to the Foot Feature Principle, given below.

102. Foot Feature Principle
The increment of the mother category's FOOT feature is the unification of the increments of the daughter categories' FOOT features.
(Gazdar and Pullum (1982:35))

This principle ensures, among other things, that all foot features of daughter nodes will also be features of the mother node.

The feature null is explicitly specified in the syntactic rule on the righthand side of the arrow in VPD metarule (98). Thus, it appears to be one of the foot features which does not obey the Foot Feature Principle. Implicit in Gazdar and Pullum's claim that foot features specified in syntactic rules "cannot be copied or otherwise tampered with in the feature instantiation mapping" is the assumption that the default value for features such as null is minus. All VP's will be [-null] by default, except the VP's specified as [+null] (i.e. VP's introduced by the VPD metarule). Thus, the VP's in the basic rules for S-Adverbs and quantifiers will be [-null] by default. Since the VP immediately following the S-Adverb or Q is [-null], rule 99 cannot rewrite this VP as the empty string e. The ungrammaticality of sentences such as 90, 92, and 93 is thus accounted for.

The contrast between 93 and 94 is readily accounted for if we assume, as do Postal (1974), Maling (1976), and Sag (1978), that sentences such as 94 have at least one structure in which the Q and Pro form an NP and that sentences such as 93 only have a structure in which Q is part of the VP. This difference in structure is supported by the parenthetical test for constituent structure. Postal (1974) and Maling (1976) both note the following contrasts.

103. Malcolm proved them all, $\left\{\begin{array}{l}\text{don't forget,}\\ \text{he claimed}\end{array}\right\}$ to be vicious criminals.

104. *Malcolm proved the soldiers all, $\left\{\begin{array}{l}\text{don't forget,}\\ \text{he claimed}\end{array}\right\}$ to be vicious criminals.

I will assume the basic rule in 105, as well as the rule which 'Chomsky-adjoins' Q's to VP's.

105. [NP -> PRO Q]

The sentence in 94 involves Subject-Auxiliary Inversion which Gazdar et al. (1982) handle with the following metarule:

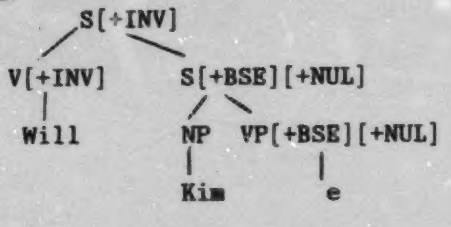
```
106.  SAI:     [VP -> V     VP],λ P[V'(V'(P))] ==>
               [+FIN]       [a]
               [+AUX]

               [S  ->  V    S],   V'(S')
               [+INV]       [a]
```

They claim that "the VPD metarule ... feeds the SAI metarule" (p. 611) and that the sentence in 107 will be assigned the structure in 108.

107. Will Kim?

108.



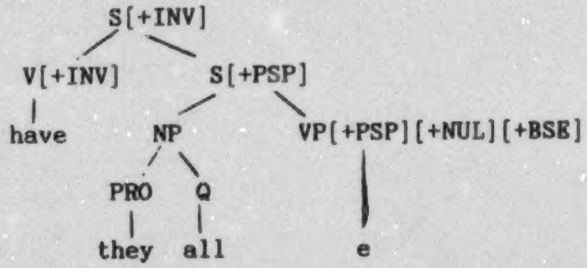
The output of the VPD metarule serves as input to SAI to allow the rule on the righthand side of the arrow in 109.

109. [VP -> V VP] ==> [S -> V S]

The tree in 108 makes it clear that Gazdar et al. are assuming that <u>null</u> is a head feature and that [+null] appears as a feature of V in 108 by virtue of the Head Feature Convention. But, as stated earlier, assuming that null is a head feature will lead to the generation of ungrammatical sentences.

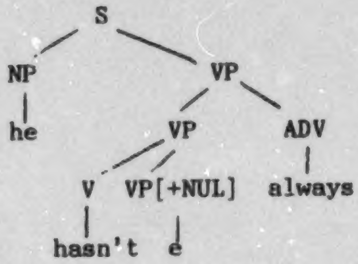
Obviously some other way of accounting for such sentences is needed. It is necessary to somehow specify that the VP in such structures may be [+null] without allowing for the generation of ungrammatical sentences. It is not obvious how this should be done. What is important for our purposes is that whatever means is used to account for such questions will also account for questions such as 93, assuming that the Q and PRO are both dominated by NP as in 110.

110.

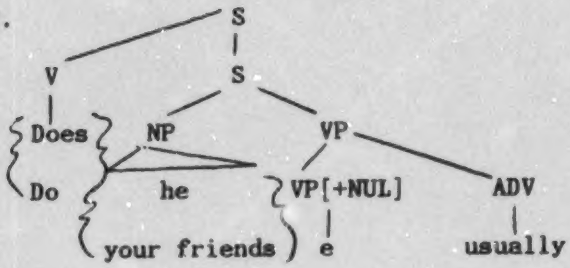

S[+INV]
V[+INV]
S[+PSP]
have
NP
VP[+PSP][+NUL][+BSE]
PRO
Q
they
all
e

Because frequency adverbs can be 'Chomsky-adjoined' to the right of VP's, the grammaticality of the examples in 95-97 is correctly predicted. They will be assigned the following structures.

111.


S
NP
VP
he
VP
ADV
V
VP[+NUL]
always
hasn't
e

112.


S
V
S
Does
Do
NP
VP
he
your friends
VP[+NUL]
ADV
e
usually

Sentences such as 113 below, with temporal adverbs will be given a similar structure.

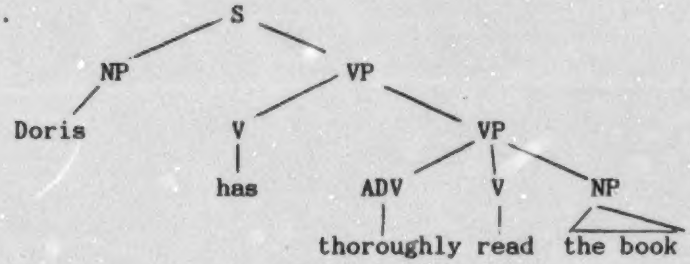
113. John will go to class today, but he won't tomorrow.

Since these adverbs are also introduced as right sisters of VP (cf. chapter 4), the grammaticality of such sentences is predicted.

The ungrammaticality of sentences such as 91 in which a VP adverb apparently precedes a VP deletion site is accounted for, because VP adverbs never immediately precede a VP. The structure for a sentence such as 114 will be 115. If the lower VP were 'deleted', the VP adverb would also have to be 'deleted'.

114. Doris has thoroughly read the book.

115.



It has been shown in this section that the facts about quantifier and adverb 'stranding' before VP deletion sites are readily accounted for given our assumptions about constituent structure and the treatment of VPD in Gazdar et al. (1982).

4.0 Advantages of the structural analysis

The structual analysis which has been presented to account for adverb and quantifier stranding facts is preferable to previous analyses for several reasons:

i. The identical 'behavior' of sentential adverbs and quantifiers before VP movement and deletion sites is explained. Quantifiers and sentential adverbs cannot precede VP extraction sites because they are 'Chomsky-adjoined' to VP's.

ii. The contrast between sentences such as 100 and 101 follows from the difference in structures these questions may have.

116. *Did the men all?
117. Did they all?

iii. The grammaticality of sentences in which frequency adverbs immediately precede VP extraction sites is accounted for.

iv. I have not discussed quantifiers and adverbs before NP extraction sites. However, assuming that quantifiers (Q) and sentential adverbs (S-ADV) are 'Chomsky-adjoined' to NP's, the ungrammaticality of sentences in which a quantifier or sentential adverb immediately precedes an NP extraction site is accounted for in the same way as the ungrammaticality of sentences in which a quantifier or sentential adverb immediately precedes a VP extraction site.

FOOTNOTES
SECTION 6

1. It might be claimed that 36 cannot be an example of gapping, since it is sometimes assumed that only two remnants may be left behind by gapping. However, there are other examples in which three constituents remain. Sag (1977:144) points out cases where the gapped clause contains three remnants (NP-PP-PP), as in his examples repeated below.
   i. Peter talked to his boss on Tuesday, and Betsy ∅ to her supervisor on Wednesday.
   ii. John talked to his supervisor about this thesis, and Erich ∅ to the dean about departmental policies.

BIBLIOGRAPHY

Bach, E. (1980), "Tenses and adverbs as functions on verb phrases," in C. Rohrer (ed.), Time, Tenses and Quantifiers, Max Niemeyer Verlag.

Baker, C.L. (1971), "Stress level and auxiliary behavior in English," Linguistic Inquiry 2, 167-181.

Baker, C.L. (1981), "Auxiliary - adverb word order," Linguistic Inquiry 12, 309-315.

Baltin, M. (1982), "A landing site theory of movement rules," Linguistic Inquiry 13, 1-38.

Bellert, I. (1977), "On semantic and distributional properties of sentential adverbs," Linguistic Inquiry 8, 337-350.

Bowers, J.S. (1969), "Adjectives and adverbs in English," Indiana University Linguistics Club.

Chomsky, N. (1970), "Remarks on nominalization," in R.A. Jacobs and P.S. Rosenbaum (eds.), Readings in English Transformational Grammar, Waltham, Mass., Ginn & Co., 184-221.

Dowty, D.R. (1978), "Governed transformations as lexical rules in a Montague grammar," Linguistic Inquiry 9, 393-426.

Dowty, D.R. (1979), Word Meaning and Montague Grammar, Dordrecht, D. Reidel.

Dowty, D.R. (1980), "Sentence adverbs and VP-adverbs revisited," paper presented at the annual meeting of the Linguistic Society of America.

Ernst, T. (1983), "More on adverbs and stressed auxiliaries," Linguistic Inquiry 14, 542-549.

Gazdar, G. (1981), "Unbounded dependencies and coordinate structure," Linguistic Inquiry 12, 155-186.

Gazdar, G. (1982), "Phrase structure grammar," in P. Jacobson and G. Pullum (eds.), The Nature of Syntactic Representation, D. Reidel, Dordrecht, 131-186.

Gazdar, G. and G. Pullum (1981), "Subcategorization, constituent order and the notion 'head'," in M. Moortgat, H.V.D. Hulst and T. Hoekstra (eds.), The Scope of Lexical Rules, Foris Publications, Dordrecht, 107-123.

Gazdar, G. and G. Pullum (1982), "GPSG: A theoretical synopsis," Indiana University Linguistics Club.

Gazdar, G., G. Pullum, and I. Sag (1982), "Auxiliaries and related phenomena in a restrictive theory of grammar," Language 58, 591-638.

Gazdar, G. and I. Sag (1981), "Passive and reflexives in phrase structure grammar," in J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Formal Methods in the Study of Language*, Mathematical Centre Tracts, Amsterdam, 131-152.

Greenbaum, S. (1969), *Studies in English Adverbial Usage*, Miami linguistic series, no. 5, Miami, University of Miami Press.

Jackendoff, R. (1972), *Semantic Interpretation in Generative Grammar*, Cambridge, Mass., MIT Press.

Jackendoff, R. (1977), *$\bar{X}$ Syntax: A Study of Phrase Structure*, Cambridge, Mass., MIT Press.

Jacobson, P. (1982), "Evidence for gaps," in P. Jacobson and G. Pullum (eds.), *The Nature of Syntactic Representation*, D. Reidel, Dordrecht, 187-228.

Jacobson, S. (1964), *Adverbial Positions in English*, Uppsala Dissertation, Stockholm: AB Studentbok.

Keyser, S.J. (1968), "Review of Sven Jacobson: Adverbial positions in English," Language 44, 357-374.

Klein, E. and I. Sag (1981), "Semantic type and control," paper presented at the annual meeting of the Linguistic Society of America.

Lakoff, G. (1965), *On the Nature of Syntactic Irregularity*, Mathematical Linguistics and Automatic Translation, Report NSF-16, Aiken Computation Laboratory of Harvard University.

Lakoff, G. (1970), "Adverbs and opacity: A reply to Stalnaker,"

Maling, J. (1976), "Notes on quantifier-postposing," Linguistic Inquiry 7, 708-718.

Postal, P. (1974), *On Raising*, Cambridge, Mass., MIT Press.

Prince, E. and G. Prince (1980), "On differentiating between Topicalization, Focus-Movement, and Yiddish-Movement," presented at the annual meeting of the Linguistic Society of America.

Pullum, G. (1982), "Syncategorematicity and English infinitival *to*," Glossa 16, 181-215.

Radford, A. (1981), Transformational Syntax, Cambridge, Cambridge University Press.

Sag, I. (1977), Deletion and Logical Form, Indiana University Linguistics Club.

Sag, I. (1978), "Floated quantifiers, adverbs, and extraction sites," Linguistic Inquiry 9, 145-150.

Sag, I. (1980), "A further note on floated quantifiers, adverbs, and extraction sites," Linguistic Inquiry 11, 255-257.

Sag, I. (1982), "Coordination, extraction, and Generalized Phrase Structure Grammar," Linguistic Inquiry 13, 329-335.

Schreiber, P. A. (1971), "Some constraints on the formation of English sentence adverbs," Linguistic Inquiry 2, 83-101.

Schreiber, P. A. (1972), "Style disjuncts and the performative analysis," Linguistic Inquiry 3, 321-47.

OSU WPL 31 (1985) 64-121

Syntactic Conditions on Two Types of English Cliticizations in GPSG

Annette Sue Bissantz

0. Introduction

Much of the recent work in morphosyntax[1] has focused on characterizing the distinctive properties of clitic elements and the grammatical role of cliticization rules in the languages of the world. Special emphasis has been placed on distinguishing clitic elements from other types of bound morphemes, devising typologies for clitics, and locating rules of cliticization within the grammar as a whole. Though not a necessary feature of such studies, the syntactic framework most often used has been some form of transformational grammar. In this thesis I will look at the phenomenon of cliticization from the point of view of a relatively new theory of syntax, that of Generalized Phrase Structure Grammar[2]. In particular I will examine two forms of English cliticization, Auxiliary Reduction and Complementizer Contraction, which have not received an adequate treatment in transformational grammars and show how they can be accounted for easily and elegantly within the GPSG framework. In a more general vein, I will also show that the nature of syntactic rules in GPSG in part predicts the existence of a separate component within the grammar for cliticization rules; a division independently argued for by many other researchers.

The two types of cliticization I will be interested in here are Auxiliary Reduction (AR) and Complementizer Contraction (CC). AR is responsible for alternations such as the one in (0.1), CC for those like (0.2):

0.1) a. Pita is almost done.
     b. Pita's almost done.
0.2) a. Pita wants to get done.
     b. Pita wansta get done.

These two rules in particular were chosen for study precisely because they have been the subject of so much discussion in recent linguistic literature. In all the debate surrounding these constructions one can isolate at least two separate issues: 1) what would be the best way of stating the conditions under which AR and CC take place and 2) how should these rules be incorporated into a grammar of English. While some theorists have claimed that the application of AR and CC is dependent upon stress levels in candidate sentences, many others have argued that syntactic structure is the primary determining factor. Furthermore, even within the latter group there has been a great deal of disagreement over how, precisely, to characterize this dependency. Similarly, the rules of AR and CC themselves have been treated differently by different researchers--being sometimes included in the phonology, sometimes in the morphology, and sometimes in the syntax of the language. In the following sections I will attempt to deal with both of these issues.

In section one I will justify my claim that AR and CC are rules of cliticization rather than, for example, simply the result of phonological reductions or affixation processes. I will also show that low stress levels do not guarantee the applicability of these rules and present preliminary arguments in favor of separating them from other rules in the grammar. In section two I present some of the more well-known formulations of AR and CC and the types of data each can and can not account for. While the main purpose of these discussions is to define the problem at hand, it should be noted that a theory which can deal with these facts in a simple and elegant way would represent a significant improvement over these alternate proposals. Section three contains a brief summary of the basic tenets of GPSG and shows how the correct generalizations about the types of syntactic structures that allow AR and CC fall out automatically in a GPSG treatment of English. This approach is particularly satisfying in that it provides a straightforward account of the dialect variations found with AR and CC. Section three also shows how assuming a GPSG syntax strengthens the conclusions reached in section one concerning the location of cliticization rules in the grammar.

1. AR and CC as Cliticizations.

A basic claim of this thesis is that AR and CC are, indeed, rules of cliticization: that is, rules which result in certain free morphemes being realized as bound dependents of other morphemes in a sentence. I wish to make a distinction here between the actual cliticization operation itself and any possible phonological consequences of that operation. As the latter are frequently idiosyncratic, the phonological form of the clitic or the clitic and its host (i.e. the morpheme on which it is dependent) will often have to be specifically listed in the grammar in much the same way as irregular past tense or plural forms. Before I present arguments in favor of this particular view of AR and CC we will need to know a bit more about the nature of clitics, their classifications and associated properties; this is discussed in the following section.

1.1 Clitics and Clitic Typologies.

Clitics are a type of bound morpheme found in many languages. They are unusual in that they act in some respects like words and in other respects like affixes, sharing certain properties with each. They are distinct from words in that they cannot usually bear stress and are phonologically dependent on a 'host' element. They can be distinguished from affixes in that they attach to already formed words rather than to roots or stems to make words, they do not necessarily have a close semantic relationship to their host word and, unlike some derivational affixes, they never affect the lexical category of their host[3]. While these are useful criteria for separating clitics as a group from words and affixes, they do not give any insight into the possible subclasses of clitic elements themselves. Many such subgroupings have been proposed. Nida (1946) divides clitics into two groups: those with alternate free forms and those without alternate free forms. Other classifications have focussed more on positioning, with many scholars[4] drawing a distinction between verbal clitics on the one hand and second position (or 2P) clitics on the other. Verbal clitics, as the name implies, attach only to verbs; they also tend to occupy a different position than their free standing counterparts. The object pronoun clitics found in many Romance languages would qualify as verbal clitics. Examples are given below from Spanish and French:

1.1) a. Veo *el libro*
see-lsg the book
'I see the book'
b. *Lo* veo
it see-lsg
'I see it'
1.2) a. Je vois *Jean*
I see-lsg IO Jean
'I see Jean'
b. Je *le* vois
I him see-lsg
'I see him'

2P clitics are typically much freer with regard to the category of potential hosts; frequently they will attach to anything that can occur in first position in a sentence. "First postition" is open to different interpretations in different languages, it could refer to the first *word* or it could mean after the first *constituent*. Klavans (1980) cites the following examples from Ngiyambaa, a language in which both interpretations are allowed. The clitic involved here is the second person nominal marker ('=' indicates the clitic boundary):

1.3) a. nadhay=ndu guya dha-yi
tasty =2NOM fish eat-past
'You ate a tasty fish'
b. nadhay guya=ndu dha-yi
tasty fish=2NOM eat-past
'You ate a tasty fish'

Similar situations attain in other languages, such as Serbo-Croatian, as well.

Zwicky (1977) represents one of the first comprehensive clitic typologies, attempting to take into account all of the factors mentioned above: i.e. host preferences, positioning and existence or lack of corresponding free forms. He divides clitics into three distinct groups-- *simple clitics*, *special clitics* and *bound words*-- on the basis of these properties. Zwicky defines a simple clitic as a phonologically reduced version of a free morpheme which becomes subordinate to a neighboring word. These reduced forms occupy the same position in the sentence as their corresponding full forms and so do not exhibit any "special" syntax. To illustrate, Zwicky cites the following example of object pronoun reduction in English:

1.4) a. He sees her
b. [hì síz hr̀] full
c. [hì sìzr̩]' reduced

The pronunciation in (1.4c) is a casual version of the sentence in (1.4a) and Zwicky notes that simple clitics are usually associated with particular speech styles or speeds.

Special clitics differ from simple clitics in two important ways. First, special clitics occupy a different position in sentence structure than non-clitic elements with the same function. So for example, in the French and Spanish sentences in (1.1) and (1.2), the clitic object pronoun *precedes* the verb while non-clitic object NP's normally *follow* the verb forming a constituent of their own; if instead the verbs were non-finite the clitic would attach to the end of the verb. Second, there is not necessarily a close phonological relationship between a special clitic and any related free form it may have (cf. the Spanish lo veo a el 'I see him', in which el is the free pronominal counterpart to the clitic pronoun lo). Thus, Zwicky concludes that these bound forms are not related to the free forms by phonological rules of any generality.

The third type of clitic in Zwicky's typology, bound words, never have free variants. While bound words attach phonologically to one word they are semantically associated with the entire constituent of which this word is a part. Since it is the constituent as a whole rather than the individual lexical item which is important, bound words can choose from a variety of lexical categories as their host. An example of a bound word would be the English possessive marker 's illustrated below:

1.5) a. The boy's hat
b. The boy who ran's hat
c. The boy who looked up's hat
d. The boy he ran to's hat

In this small number of examples alone the possessive marker is attaching to a noun, a verb, a particle and a preposition though they all are part of an NP constituent.

The problem for this approach is that clitics in many languages do not always fall into these three neat groups. Some clitics may, for example, act like bound words in some respects and like simple clitics in others. Klavans (1980) criticizes Zwicky's typology on just this point also arguing that his approach does not provide a framework in which to describe historical changes in clitic systems or capture similarities and differences between certain clitic types. In particular, Klavans charges that Zwicky's claim concerning the development of bound morphemes--that independent words are reanalyzed as clitics which are then reinterpreted as affixes--lacks motivation in some instances and is historically inaccurate in others. She further objects to the failure of Zwicky's typology to recognize similarities between clitics based on positioning. Klavans cites the example of 2P pronouns in Walpiri and 2P particles in Tagalog: the former are classified as special clitics while the latter are said to be bound words. Thus the fact that clitics seem to be drawn to certain positions in a wide range of languages is obscured.

Klavans rejects earlier typologies of clitics and clitic placement as being too simplistic and suggests that such facts can be given a unified account only by characterizing them in terms of the following five parameters:

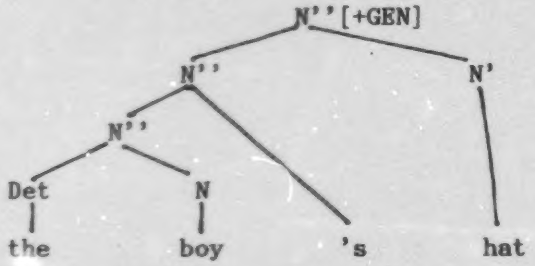
P1: Clitic Identity
P2: Domain of Cliticization
P3: Initial/Final
P4: Before/After
P5: Proclitic/Enclitic

P1 merely refers to a lexical feature [+clitic] by which clitics can be identified by cliticization rules. P2-P4 are concerned with the syntactic placement of the clitic. P2 refers to the node with respect to whose immediate constituents the syntactic position of a clitic is determined. P3 indicates whether it is the first or last immediate constituent in the domain which is relevant for placement and P4 whether the clitic attaches to the left (Before) or to the right (After) of this constituent. P5 makes explicit the phonological attachment of the clitic; if it attaches to the end of the preceding word it is enclitic, if it attaches to the beginning of the following word it is proclitic. To give an example, the possessive marker in English would have the following values for these five parameters:

P1: English possessive
P2: N''[+GEN]
P3: Initial
P4: After
P5: Enclitic

Since possession is marked on (genitive) NP's this is the domain of cliticization (P2). P3 is initial because the first constituent in NP is marked, i.e. the boy in something like the boy's hat. P4 is after because the marker follows the constituent picked out by P3 and P5 is enclitic because the marker combines phonologically with the preceding material. The tree in (6) illustrates the syntactic positioning:

1.6)



Klavans argues that this typology is superior to Zwicky's because it can capture similarities in syntactic positioning quite straightforwardly and is superior to typologies based soley on a verbal vs 2P distinction in that it allows for a greater range of clitic positions (i.e. eight possible locations per domain). This last feature is, in fact, precisely the problem with her approach; her system is simply too unrestricted. While Klavans claims to have substantiated each of the eight clitic positions (p. 138), the examples she gives are not all from the same domain. As can readily be seen on closer examination it would be impossible to substantiate each of the positions for every domain since some combinations of parameters are nonsensical. Take, for example, the following two parameter combinations:

1.7) a. P1: --
P2: S
P3: Initial
P4: Before
P5: Enclitic

b. P1: --
P2: S
P3: Final
P4: After
P5: Proclitic

Each of the sets in (1.7) would require the clitic to attach to something outside of its own S; neither of these clitic types have been convincingly attested[5].

Of the clitic positions Klavans does attempt to support, some are based on less than persuasive evidence -- a case in point being her categorization of Old Indic Preverbs. Klavans argues for the positioning of Old Indic Preverbs by appealing to the analyisis of Proto-Indo-European preverbs in Anderson (1979). As she herself admits (p. 138) the evidence is only suggestive, it is far from conclusive. Another problem with this typology is that it predicts that every clitic position is just as likely to occur as any other; it gives no explanation of why some positions turn up in language after language in many different families while some positions don't seem to turn up at all. Thus Klavans' analysis is no more informative on this point than Zwicky's and certainly cannot be considered superior to it. Furthermore, Klavens' analysis fails to distinguish in a systematic way between clitics which have free standing counterparts and clitics that don't, overlooking an obvious and, for our purposes, important typological difference. In any case, since our main interest here is not so much with the range of possible positions for clitics (or for that matter with characterizing their historical development), but with their associated properties, Zwicky's system will ultimately be of much more use. Where clitic positioning is relevant we will rely on the standard verbal vs. 2P distinction. In the next section we will take a closer look at the rules of AR and CC and see how reduced auxiliaries and complementizers fit into the framnework assumed above.

1.2 AR and CC.

As we have seen, AR is an optional process by which finite forms of certain auxiliary verbs[6] become dependent on neighboring material. In most dialects reduced auxiliaries show a low degree of selection with regard to the category of the lexical items they attach to. Instead, what seems to be important is the category of the constituent this lexical item is part of[7], as we can see from the examples below:

1.8) a. Pita's a cat.
b. He's a cat.
c. The cat Mary painted red's named Pita.
d. The cat Mary hit's named Pita.
e. The cat Mary talked to's named Pita.
f. The cat Mary fed yesterday's named Pita.

Since reduced auxiliaries have alternate free forms that occupy the same position in the sentence, in Zwicky's typology they would be classified as simple clitics rather than special clitics or bound words. As such we would expect them to display the same type of behavior as other simple clitics and, as we shall see a bit later, this is indeed the case.

While some researchers, most notably Bresnan (1971), have argued that reduced auxiliaries must be treated as proclitic to following material in order to account for sentences in which AR is blocked, most clitic analyses have viewed AR as a rule of enclisis. This, plus the facts that AR applies to finite verb forms and finite verb forms usually follow subject NP's in English, gives reduced auxiliaries something of the appearance of 2P clitics. In fact, this very property is exploited in an interesting discussion of possible causes for dialect variation in sentences involving AR presented in Kaisse (1983b). Notice, however, that reduced auxiliaries cannot in general be treated as 2P clitics because of the existence of sentences like (1.9b):

1.9) a. John unfortunately is not going.
     b. John unfortunately's not going.

In this case the adverb unfortunately is occupying second position and the auxiliary is is reducing onto it. Thus, while it is frequently the case that reduced auxiliaries are 2P, it is not necessarily true. We will return to the question of proclitic versus enclitic treatments in section two.

CC, like AR, is also an optional rule which results in the reduction of a free morpheme (infinitival to) onto preceding material. Since contracted to's, like reduced auxiliaries, have alternate free forms which occupy the same syntactic position, they too would be considered simple clitics in Zwicky's system. Additional examples of CC are given in (1.10):

1.10) a. They wanna be in pictures.
      b. They hafta be in pictures.
      c. They usta be in pictures.
      d. They oughta be in pictures.
      e. They gotta be in pictures.
      f. They're sposta be in pictures.

In most accounts of CC it would be possible to view contracted to as a verbal clitic since it is either assumed that this reduction can only occur with a few, lexically specified verbs[8] (hence the common name Wanna-Contraction) or with the class of verbs as a whole. The one exception to this is Jacobson (1982) who claims that to can cliticize onto both verbs and adjectives, the particular lexical item involved being irrelevant.

Jacobson bases this claim on sentences with reduced vowels, like those in (1.11)-(1.13), which she says are grammatical for some speakers:

1.11) a. I want tə.
      b. He wants tə.
1.12) John seems tə.
1.13) John is expected tə.

She also argues that, even for speakers who disfavor (1.11)-(1.13) there is a sharp contrast between those sentences and ones in which the item preceding the to is not a verb or adjective:

1.14) *I want Sam tə.
1.15) *I persuaded Sam tə.
1.16) *I want very much tə.

All of the sentences in (1.11)-(1.16) seem equally awkward to me, those in (1.14)-(1.16) no more so than the others. But even if there are speakers who share these judgments, Jacobson's conclusions are not warranted. The problem lies in distinguishing actual cliticization from simple phonological vowel reduction. There are two types of evidence in favor of the latter analysis for at least some of Jacobson's examples. One is the critical interplay between her to reduction rule and stress--a known factor in phonological reductions. The second is that reduction of to to tə is possible even in contexts in which CC is not allowed. An example of this type would be (1.17b):

1.17) a. Who does Pita want tə kiss you?
b. *Who does Pita wanna kiss you?

where (1.17b) is not possible in most dialects. Another example would be Jacobson's sentence (75) given below as (1.18):

1.18) Tə run is no fun.

Since CC involves leftward cliticization the tə in (1.18) could not be a result of the same rule. To account for these sentences Jacobson must posit a second, otherwise unmotivated cliticization rule to perform the same function as well-founded phonological rules. Thus it is clear that Jacobson is attempting to account for too wide a range of phenomena with her rules.

For the purposes of this work I will adopt the view that CC applies to the class of verbs (all verbs and only verbs) with unpredictable phonological effects in some cases and predictable effects in others. The sentences in (1.11)-(1.16) will be attributed to the operation or failure of a phonological reduction rule rather than cliticization. This treatment will allow us to capture the contrast between sentences like (1.11) and (1.12), which some speakers reject, and sentences like (1.19) and (1.20), which they find completely acceptable:

1.19) I wanna.
1.20) He wansta.

Notice, however, that this definition of the domain of CC is not a necessary feature of my analysis. If future evidence persuasively argues in favor of one of the other proposed domains for CC the change can easily be effected using subcategorization. As things stand this slight complication does not seem to be needed. As noted above, this view of CC is consistent with the claim that reduced to is a verbal clitic.

1.2.1 AR, CC and Phonology.

The contrast between (1.17a) and (1.17b) noted above also argues against the claim (suggested by Lakoff (1970), among others) that CC is conditioned by low stress and thus is a phonologically determined rule. Since Jacobson's to reduction rule, which is conditioned by stress, can apply to produce (1.17a) if CC were also stress dependent we would expect it to be able to apply here as well. The fact that CC is ungrammatical in (1.17b) shows that something else is going on in these sentences. A similar argument can be made for AR as well, as pointed out in Kaisse (1983a). While the unstressed auxiliaries in (1.21a) and (1.22a) can be phonologically reduced to [z] or [ɨz] in most dialects, they cannot be realized as a fully reduced clitic forms, i.e. without any vowel at all, in any dialect:

1.21) a. I wonder how much wine there is in the bottle.
b. *I wonder how much wine there's in the bottle.
1.22) a. John's nicer in the mornings than Harry is at night.
b. *John's nicer in the morning than Harry's at night.

Thus, lack of stress cannot be the determining factor in the operation of AR or CC. The strong correlation that is found between clitics and stresslessness in many languages can be accounted for in other ways; for example, by having rules remove stress from cliticized elements or even by ordering stress assignment rules after cliticization and having them fail to operate on clitics. The point is, we need not and cannot assume that stress is what conditions the operation of the rules under discussion.

In addition to the claim that AR and CC are phonologically conditioned, it has also been suggested that AR and CC are themselves phonological reduction rules. Due to the highly idiosyncratic effects of these rules, however, such an analysis is unworkable as well. As Kaisse (1983a) points out, the phonological rules that would be needed to derive reduced auxiliaries from their full counterparts are either not productive rules of English at all or not productive at all the speech rates which permit AR:

1.23) a. is [s], [z], [əz] (or [ɨz])
b. are [r̩], [r]
c. am [m̩], [m]
d. has [s], [z], [əz] (or [ɨz])
e. have [v], [əv]
f. had [d], [əd]
g. will [l̩], [l]
h. would [d], [əd]

For example, Kaisse notes that English has no regular rule of [w] deletion, which would be needed in a phonological derivation of (1.23g) or (1.23h). Also, while there are productive rules to delete [h] when it occurs before an unstressed vowel, they apply only in rapid speech. Since AR is possible even at relatively slow speech rates, forms like (1.23d-f) could not be generated. Similarly, even though full vowels can reduce to schwa at all speech rates, the rules which delete schwa entirely are also restricted to fast speech; thus none of the vowelless alternates in (1.23) could be derived at a slower rate either.

Finally, if we examine the reduced alternates of is and has given in (1.23a) and (1.23d) respectively, we notice that they are suspiciously similar to the various allomorphs of the plural, third person singular and possessive morphemes both in form and distribution: only [əz]/[ɨz] can occur after stridents while [s] occurs after voiceless non-stridents and [z] after voiced non-stridents[9]. The most general way of accounting for these facts would be to allow the rules which determine the distribution of the allomorphs of these other morphemes to also determine the distribution of the reduced forms of is and has. Since in most recent theories of grammatical organization rules of this type precede phonology proper, the rules which determine when an auxiliary can be realized as its reduced form (as opposed to its full form) must also precede the phonological component and, therefore, must be of a distinct type.

There are similar arguments against treating CC as a phonological reduction, as well. First of all, as with AR, the phonological rules that would be needed to derive CC forms from their full counterparts are not all fully productive. For example, in order to derive the reduced sentence in (1.24b) from its full counterpart in (1.24a):

1.24) a. I want to finish.
b. I wanna finish.

we would need both a degemination rule and a nasal assimilation rule with at least one operating across a word boundary. While rules operating across word boundaries are common in English they are generally restricted to fast or casual speech. Forms like (1.24), however, are perfectly acceptable even in slow, careful styles. Even more disturbing is the fact that there is really no well defined set for these rules to apply to. AR, at least, can be restricted to the class of auxiliary verbs, though not all auxiliary verbs are affected. The verbs which undergo radical phonological changes as a result of CC, however, have no other common properties to set them apart from other verbs. Thus there would be no general way of preventing the derivation of sentences like (1.26b) alongside (1.25b):

1.25) a. I want to live.
b. I wanna live.
1.26) a. I hunt to live.
b. *I hunna live.

In my speech want and hunt differ primarily only in initial consonant, thus there would be no phonological grounds on which to exclude (1.26b). It is obvious, then, that the relationship between want to and wanna needs to be stipulated rather than derived. Since this type of "spelling out" rule is typically found in the morphological component, e.g. take + past tense = took, and the morphological component is typically ordered before phonology, we again have an argument for ordering the rules governing the distribution of full (versus reduced) forms before phonology. Notice that these facts are perfectly consistent with the view that AR and CC belong to a separate component of the grammar reserved for cliticization and ordered between syntax and morphology, as argued for in much of the recent literature (see references, fn. 1).

1.2.2 AR, CC and Morphology.

Another possibility that should be considered here is that reduced sentences are not derived via productive rules at all but, rather, hosts bearing reduced elements are listed separately in the lexicon and assigned the appropriate distribution (e.g. wanna alongside want, John's alongside John). While such an approach to AR is totally unworkable, it is at least plausible in the case of CC. Since the reduced alternates of auxiliaries like is and has appear quite freely with preceding NP's no matter what their composition, it would be impossible to limit the number of different constructions in which they occur. Thus we would either have to list an infinite number of otherwise perfectly regular phrases separately in the lexicon or allow the word that bears the reduced auxiliary, no matter how deeply embedded it may be, to determine the type of matrix VP that is allowed. This is clearly absurd. On the other hand, since reduced to has a much more restricted distribution than reduced auxiliaries, occurring only with verbs that can take an infinitival complement, it would be relatively simple to separately list forms with reduced to and forms without reduced to for each such verb. The forms with reduced to would differ from those without only in that they subcategorize for bare infinitive complements rather than overt infinitive complements.

The problem with this approach, aside from the distributional peculiarities and the redundancy of listing both forms, is that forms like *wanna*, *gotta*, etc. do not function *syntactically* like single words as we would expect if they had separate lexical entries. They do not undergo any type of derivation or inflection, nor are they operated on as a unit by any syntactic rules. In fact, sentences in which they *are* treated as a unit are judged to be ungrammatical. For example, compare (1.27b) with (1.27c):

1.27) a. John is supposed to drive to Cleveland and Mary is supposed to fly to Toledo.
b. John is sposta drive to Cleveland and Mary is sposta fly to Toledo.
c. *John is sposta drive to Cleveland and Mary is fly to Toledo.

If in fact *sposta* were a separate lexical item we would expect it to undergo gapping, just like any other verb:

1.28) John will drive from Cleveland to Toledo and Mary will __ from Toledo to Akron.

The fact that (1.27c) is ungrammatical shows that sposta is not a syntactic unit but merely a phonological one. Thus this type of morphological treatment of AR and CC, the lexical approach, cannot work.

A second type of morphological treatment which has been argued for is the view that AR and CC involve affixation rather than cliticization. There are, however, a number of reasons for not believing this to be the case. One such reason is that reduced auxiliaries and contracted *to* have more properties in common with clitics than they do with affixes. Zwicky and Pullum (1982) present the following criteria for distinguishing between simple clitics and affixes (Z&P. p.3):

1.29) a. Clitics exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.
b. Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.
c. Morphophonological idiosyncracies are more characteristic of affixed words than of clitic groups.
d. Semantic idiosyncracies are also more characteristic of affixed words than of clitic groups.

If we measure the results of AR against the principles in (1.29) we see, as Zwicky and Pullum themselves point out, that reduced auxiliaries are almost a paradigm example of simple clitics.

Examining just the sentences given in (1.8) above we find examples of an auxiliary verb reducing onto a noun, a pronoun, an adjective, a verb, a partical, and an adverb. From this we can see that, though there may be general restrictions on the preceding constituent in some dialects, the category membership of the word the auxiliary actually attaches to is not important; reduced auxiliaries do indeed exhibit a low degree of selection. Furthermore, unlike affixes, there are no cases in which a particular lexical item idiosyncratically blocks the application of AR.

There are cases in which AR is disfavored (not blocked) with particular lexical items, but these are for perfectly straightforward phonological reasons. The phonological effects of combining a reduced auxiliary with its host are also perfectly straightforward. While irregular plural or past tense forms are quite common, the phonological variations in reduced auxiliaries are fixed and predictable from the phonological and morphological properties of the host. Finally, there are no cases in which the semantic contribution of the reduced auxiliaries is in any way different from the semantic contribution of the corresponding full form.

Though contracted to's do not fare quite so well with respect to the criteria in (1.29) they do, nonetheless, have some distinctly non-affixal properties. The fact that CC does not allow a wide range of categories to act as host does not necessarily reflect on its status as a cliticization rule since, as we saw above, a large number of clitics are restricted to verbal hosts. This is just one way in which contracted to's are less like simple clitics than reduced auxiliaries are. Since, by our definition, CC will reduce to onto any verb we do not have arbitrary gaps in the set of possible combinations. We do, however, have morphophonological idiosyncracies in a few of these combinations. Notice though that the total number of such idiosyncracies is much lower than for verbal paradigms. Notice also that such irregularities can occasionally be found in known clitic groups as well (Spanish le lo -> se lo), they are merely less frequent. As with reduced auxiliaries, the semantics of contracted to is entirely compositional. In sum then, contracted to does not exhibit any behavior that cannot be attributed to some type of clitic (though not always simple clitics) though it does lack certain properties frequently found in affixes.

There are other reasons for rejecting an affixal analysis of AR and CC as well. For one thing, treating these rules as affixation would greatly complicate the morphology of English. In addition to paradigms like (1.30):

1.30) a. I want
b. you want
c. he, she, it wants
etc.

we would have ones like the following:

1.31) a. I wanna
b. you wanna
c. he, she, it wansta
etc.

This would be true for every verb that underwent CC (i.e. for every verb in the language that takes an infinitival complement). We would also have to somehow insure that such verb forms are followed by verb phrases beginning with bare infinitives. This would be a novel situation in that it would be the affix subcategorizing the following material rather than the verb itself. The situation with AR would be even worse since reduced auxiliaries can attach to elements from so many different categories; we would in effect be creating a group of affixes that can attach to almost any word in the language but are semantically associated with the entire sentence. Again, this is clearly absurd. Consider also the fact that affixed words can be treated as units by syntactic rules unlike the products of AR and CC, as we saw above. In the case of AR, as Zwicky and Pullum point out, such a syntactic rule would be almost inconceivable.

Perhaps the most persuasive reason for rejecting an affixation analysis of AR and CC, however, is the fact that both operations are sensitive to aspects of the sentence other than just the word they are attaching to. Compare the following pairs:

1.32) a. Who does Pita wanna see?
b. *Who does Pita wanna see you?
1.33) a. Who's going?
b. *Who's?

While all the sentences in (1.32) and (1.33) are grammatical with their corresponding full forms, only the (a) sentences allow reduction. This is an important difference between affixation rules and AR and CC; while the conditions governing the combining of affixes with their stems are purely morphological and lexical, those governing the application of AR and CC seem to be syntactic in nature. This argues in favor of a separate, non-affixal analysis of AR and CC. Thus, all things considered, the clitic analysis of reduced auxiliaries and contracted to is more strongly supported by the evidence and we can conclude that AR and CC are, in fact, rules of cliticization rather than affixation or phonological reduction.

1.3 Cliticization and Syntax.

In the preceding sections it was argued that AR and CC are conditioned by syntactic structure rather than by phonological, morphological or lexical considerations. It should be noted that this is very different from the claim that AR and CC are themselves syntactic rules. In fact, contra Bresnan (1971), there does not seem to be very much evidence for the claim that cliticization rules belong in the syntactic component of the grammar. Notice, first of all, that there are no syntactic rules whose operation depends on the application of a cliticization rule. Nor, as we will see in section two, are there any syntactic operations that are bled by a cliticization rule either. Furthermore, cliticization rules are of a very different type than other syntactic rules dealing, as they do, with units smaller than words rather than entire words and phrases. This is all consistent with the view that rules like AR and CC form their own component in the grammar, one dealing with the production of phonological words rather than syntactic words. While this is a much more restricted model of grammar in that it severely limits the range of possible rule interactions, it is in no way predicted by current transformational frameworks. In section three I will show that, given a GPSG syntax, this type of organization falls out automatically; thus supporting a conclusion reached on independent grounds by many others (see references fn. 1).

1.4 Conclusions.

In the preceding discussion I have argued for the claims that 1) AR and CC are, in fact, synchronic rules of grammar, 2) that they are best analyzed as belonging to a separate component of the grammar reserved for rules of that type and 3) that the primary factor in determining the applicability of AR or CC is the syntactic structure of the candidate sentence. In the following two sections I will discuss the issues of how these conditions on syntactic structure should be formulated and what the optimal analysis shows about the grammar as a whole.

2. Previous Analyses

A number of different analyses of CC and AR have been proposed over the years with widely different views of how the cliticization process fits into the framework of a grammar. Most, if not all, of these analyses have recognized the need to refer to syntactic structure when describing the conditions under which these rules apply. These treatments can be loosely grouped into three types: those requiring some sort of explicit global reference, those involving the transformational cycle, and those appealing to some form of trace element. In what follows I will briefly review some of the more influential of these past analyses while pointing out some of the problems these treatments have had. I will return to the discussion of the place of cliticization rules in the grammar in section 3.

2.1 Global Rules.

Perhaps the best known discussion of AR and CC is the "global rule" analysis given in Lakoff (1970). Lakoff sees both AR and CC as purely phonological reductions and argues that since they are sensitive to aspects of syntactic structure they must be global rules. Lakoff bases his formulation of the conditions on AR on facts about where be can reduce first noticed by King (1970):

2.1) a. i. There's a man in the room.
ii. *I asked which men there're in the room.
b. i. It's hot.
ii. *. . . and hot it's.
c. i. You said the concert's in which park?
ii. Which park did you say the concert's in?
iii. *In which park did you say the concert's?
d. i. Kim is to leave and Sandy's to, also.
ii. *Kim is to leave and Sandy's, also.

While sentences like (ai, bi, ci, cii, and di) allow be to contract, the corresponding sentences in which Wh-Movement, Topicalization, or VP-Deletion have disturbed the complement of the auxiliary do not. Lakoff (p. 631) cites the following generalization "If there is a constituent immediately following be, and if by any transformation that constituent is deleted, then the be cannot contract." The problems with this formulation are well known. While Lakoff can account for the contrasts in (2.1), his analysis makes incorrect predictions about the grammaticality of the sentences in (2.2) and (2.3):

2.2) John's to force himself to stop.
2.3) a. Where's the library?
b. What's a global rule?
c. How fat's your cat?
d. In which city's the conference?

In his transformational framework, no matter how you order to Insertion and Equi NP Deletion the be in (2.2) would be followed by a movement or deletion site and, therefore, should not be contractable. Similarly, the sentences in (2.3) would be ruled out since they involve not only the movement of the constituent following be, but the movement of *be* itself.

Lakoff also proposes a global constraint on CC to account for the contrast in meaning between sentence pairs like the following, first noticed by Horn (cited in Lakoff (1970)):

2.4) a. Teddy, I want to succeed.
b. Teddy, I wanna succeed.

Sentence (2.4a) is ambiguious between the readings *I want Teddy to succeed* and *I want to succeed Teddy* while sentence (2.4b) can only have the second interpretation. Lakoff concludes from such sentences that CC is blocked if at any stage in the derivation an NP had intervened between the verb and *to*. The deep structures for the sentence in (2.4a) would presumably be those in (2.5):

2.5) a. I want [Teddy succeed]
b. I want [I succeed Teddy]

Notice that for this analysis to work Lakoff must explicitly order *to* Insertion *after* Equi NP Deletion (to permit contraction in (2.5b)) but *before* the rules responsible for topicalization (to block contraction in (2.5a)). This seems to be the only motivation for such an ordering.

Another problem with this analysis is that it fails to block contraction in sentences like the following, taken from Pullum and Postal(1982):

2.6) a. To regret what one does not have seems like to want.
b. ?It seems like to want to regret what one does not have.
c. *It seems like to wanna regret what one does not have.
2.7) a. I don't want anyone [who continues to want] to stop wanting.
b. *I don't want anyone [who continues to wan]na stop wanting.
2.8) a. I want to dance and to sing.
b. *I wanna dance and to sing.
2.9) a. I don't need or want to hear about it.
b. *I don't need or wanna hear about it.

Though these sentences satisfy the condition on intervening NP's, none of them allow contraction.

A revised version of Lakoff's constraint on AR is presented in Kaisse (1983a) where it is suggested (p.93) that the original condition be interpreted as in (2.10):

2.10) Auxiliary Reduction may not apply if the element following the auxiliary is not the same as the element that follows it at the stage in the derivation prior to all movements and deletions.

In addition to the sentences in (2.1), this formulation accounts for the ungrammaticality of AR in something like (2.11b) in which an element has been *inserted* following the auxiliary:

2.11) a. He is, I should think, a bit tired.
b. *He's, I should think, a bit tired.

The constraint in (2.10) does not, however, make the necessary distinction between rules like Equi and There Insertion (which do not block reduction) on the one hand and Wh-Movement and Topicalization (which do block AR) on the other. Nor does it sufficiently limit the class of hosts to those permitted in Kaisse's dialect. To remedy these defects Kaisse adopts a modified "split T" model of grammar in which "move NP" rules are distinguished from "move Wh" rules[10]. AR is then made sensitive to the level of structure resulting from the "move NP" rules and the set of possible hosts to AR restricted to NP's. The model of grammar Kaisse assumes is shown in (2.12):

2.12) (Base Rules)
↓
D-Structure
(Move NP)
↓
NP-Structure
(Move Wh)
↓
S-Structure

Control
Predication
Binding Theory
Case Marking
To Complementizer
Contraction
Identity Filters

Deletions
Stylistic Rules
Restructuring Rules
Morphological Rules
Phonological Rules

Following Pullum and Zwicky (forthcoming) cliticization rues are treated as part of a separate component, labeled here as "Restructuring Rules". Given this model, Kaisse's restriction on AR is as follows:

2.13) X NP AUX Y Z -----> 1, 2#3, 0, 4, 5
1 2 3 4 5
where 2 c-commands 3, and 4 follows 3 at NP-Structure

This says that AR is possible just in case the host is a noun phrase which c-commands the auxiliary and the element following the verb to be cliticized followed it at NP-Structure. Thus, the starred sentences in (2.1) are blocked since rules have applied to the NP-Structure which have altered the material following the verb. Presumably a sentence like (2.2) would be generated without a subject NP in the lower clause, thus contraction is possible. As it stands the rule in (2.13) also incorrectly predicts that sentences like (2.3a) and (2.3b), repeated below, are ungrammatical:

2.3) a. Where's the library?
b. What's a global rule?

To account for cases like these Kaisse includes a rider on her restriction which allows reduction onto a nonlexical item, such as a Wh-word, as long as the element following the verb meets certain criteria. This rider is given in (2.14):

2.14) In addition, if 2 is a (monosyllabic) pro-form it need not be an NP, and it suffices that X [Y: AB] not mark a movement or deletion site.

Kaisse's constraint on AR makes many other predictions as well. Thus all of the sentences in (2.15)-(2.20) will also be blocked by (2.13):

2.15) Which dog's he buying?
2.16) Not only's Louis smart, he's also a varsity rower.
2.17) On which day's John leaving?
2.18) a. Speaking tonight's a famous reporter.
b. Speaking tonight's been a famous reporter.
2.19) a. More important's her insistence on honesty.
b. More important's been her insistence on honesty.
2.20) Under this slab's buried Joan of Arc.

(examples based on Kaisse (1983a)). Sentence (2.15) is bad because, due to the application of Subject Auxiliary Inversion (SAI), the element following be at the time cliticization operates is not the same as the element which followed it at NP-Structure. Sentences (2.16) and (2.17) are rejected on two counts: SAI has applied in these sentences (triggered by various preposing rules) and the host for the clitic is not an NP. While the sentences in (2.18), (2.19) and (2.20) do not involve SAI (cf (2.18b) and (2.19b)) they still fail the NP host condition.

While the constraint in (2.13) may adequately describe Kaisse's dialect, it does so at the cost of employing an extremely powerful mechanism -- a global rule. In addition to this, there are dialects in which all of the sentences given in (2.15)-(2.20) are perfectly grammatical[11]. This poses a particularly difficult problem for Kaisse's analysis since some of these sentences violate both conditions of her constraint at the same time. Thus there would be no way of generalizing Kaisse's constraint to include this other dialect. Since an analysis which accounts for different dialects with a related set of rules is to be preferred over one which treats them with entirely separate rules, Kaisse's constraint is less than satisfactory.

Kaisse (1983b) presents a modified version of this analysis in which the condition on preceding context is altered to bring AR more in line with the behavior of similar clitics in other languages. Kaisse argues that reduced auxiliaries are second position clitics and, as such, should not be sensitive to the category membership of their host. Thus she replaces her NP host condition with the following:

2.21) An auxiliary may only cliticize onto the first word of its S.

This constraint rules out the sentences in (2.15)-(2.20) since the various preposing operations involved -- Wh-Movement, Comparative Preposing, PP Fronting, etc. -- move material into COMP and outside the domain of S. As a result, the auxiliary itself is the first element in the S in these sentences and, thus, cannot appear in reduced form. However, according to Kaisse's own article sentences like (2.22) involve Wh-Movement of the subject phrase:

2.22) Which man's going to win?

Since Wh-Movement would insert the subject phrase into the COMP node, Kaisse's analysis predicts that sentence (2.22) is also ungrammatical and for the same reason as (2.15) and (2.17). This judgment is not confirmed in any dialect studied to date including Kaisse's. Kaisse notes a similar problem with sentences like:

2.23) Jack is the man who I bet's going to win.

Since the auxiliary would be the first element in the embedded S, the structure should not permit reduction[12].

The constraint in (2.21) does, however, allow for relatively more dialect variation than the NP-host condition since the domain involved can be easily modified. Thus Kaisse can account for the fact that sentences like (2.18b), (2.19b), and (2.20) with preposed elements in COMP are perfectly grammatical for many speakers by changing the S in (2.21) to S' for these dialects. However, if Kaisse is still assuming the restriction on following context given in (2.13)[13], she cannot explain the grammaticality of the parallel sentences in (2.18a) and (2.19a) or the sentences in (2.15), (2.16) and (2.17) for these same speakers. Not only does the revised constraint in (2.21) incorrectly predict the facts of Kaisse's own dialect, it still fails to account for the judgments found in other dialects. A GPSG analysis, on the other hand, presents a unified analysis of both.

2.2 Cyclic Treatment.

Bresnan (1971) proposes making rules of cliticization such as CC and AR part of the transformational cycle rather than including them in the phonology as Lakoff did. In her analysis of CC, to can cliticize leftward onto the proper type of verb if they are adjacent during that verb's cycle. Thus sentences like (2.24a) will be allowed to undergo CC since the subject of the lower clause is removed by Equi on the want cycle, leaving it adjacent to to.

2.24) a. You want [you kiss who]
b. Who do you wanna kiss?

A sentence like (2.25b), however, will not be produced since who is moved to the front of the sentence from its position between want and to by Wh-Movement on the higher, S', cycle. Hence, since CC is presumed to be cyclic, it never gets a chance to apply:

2.25) a. You want [who kiss you]
b. *Who do you wanna kiss you?

The ambiguity contrast found in (2.4a) and (2.4b) would be parallel to this example: only one of the readings of (2.4a), that corresponding to the deep structure in (2.5b), has want and to adjacent on the want cycle; the other has an intervening NP. Notice that this approach also accounts for the ungrammaticality of (2.6c) and (2.7b). Assuming strict cyclicity, the to in these sentences would not be eligible for contraction onto the want since they are not part of want's complement. Crucially, however, it does not explain the failure of CC to apply in sentences like (2.8b) and (2.9b). In these cases the to phrase is the complement of an appropriate verb and the two are adjacent on that verb's cycle, and yet contraction does not take place.

Bresnan's cyclic analysis of AR also runs into problems. In an attempt to explain why the material following the verb should be relevant to reduction, Bresnan reanalyzes AR as a rule of procliticization in which the auxiliary is attached to the front of the next word. In this treatment the starred sentences in (2.1) would have to have structures such as the following:

2.26) a. *I asked which men there 're_ in the room.
b. *. . . and hot it 's_.
c. *In which park did you say the concert 's_?
d. *Kim is to leave and Sandy 's_, also.

These could not, however, be generated once cliticization has taken place, since the transformations involved are not defined over subparts of words.

This type of analysis fails on both syntactic and phonological grounds. As Lakoff (1972) points out, a proclitic treatment of reduced auxiliaries would be very peculiar given the fact that clitic has and is--like plurals, possessives, past tense and third person singular markers--assimilate in voicing to what precedes not what follows. In order to account for this fact Bresnan would have to posit an otherwise unmotivated word external process to perform precisely the same function as a well documented word internal process, thus missing an obvious generalization and unnecessarily complicating the phonology.

A syntactic argument against this analysis is given in Wood (1979). Wood notes that Bresnan's treatment of AR cannot account for the grammaticality of sentences like:

2.27) Herb's going and Jerome is __ too.

If AR is cyclic then it applies or fails to apply on the same cycle for each conjunct. Therefore, after the first cycle the lowest verb phrase of the left conjunct would be 'sgoing while the lowest verb phrase of the right conjunct would be going. Thus the identity condition on VP Deletion would not be met and the sentence in (2.27) could not be generated.

A final problem with this approach stems from the nature of cliticization rules in general. Clitic elements, unlike some types of affixes, do not change the category of their hosts; e.g. wh-words with clitics attached are the same category as they would be without the clitic. Since Bresnan views cliticization as a process by which elements become syntactic dependents of preceding or following elements, in order to prevent sentences like:

2.28) a. *I asked 're-which men there __ in the room.
b. *'s-In which park did you say the concert __?
c. *. . . and 's-hot it.

one would have to somehow build into *each* movement rule a clause that examines the structure of the constituent in order to detect the presence of any clitic elements and prevents the rule from applying if such an element is found. This would complicate these rules enormously. Furthermore, as we saw in section one, Bresnan's basic assumption--that cliticization rules can and should be included in the syntactic component of the grammar--is not supported by rule interaction facts.

2.3 Trace Theories.

Perhaps the most frequently appealed to type of analysis is one involving some sort of trace element. In such analyses, cliticization is possible only if traces do not appear in the relevant positions in syntactic structure. What form these traces take and how, precisely, they arise is a matter of considerable variation from theory to theory. Selkirk (1972) proposes an analysis in which traces take the form of extra word boundary markers which serve to block the destressing rules that feed various cliticizations. According to her analysis, word boundary symbols flank members of major categories in deep structure. When transformational rules move or delete elements they leave the position of these boundaries unaffected. When a moved item is adjoined elsewhere in the sentence new boundary markers are created. Selkirk also includes a convention by which redundant internal boundary symbols are deleted in the configurations W#]#]Z and Z[#[#W as long as the outermost bracket is not labeled S'. The destressing rule relevant to our concerns is Selkirk's "Monosyllabic Rule" which removes stress from monosyllabic dependents that are followed by at most one word boundary symbol followed by a word with a stressed vowel. Thus a sentence like (2.1aii) could not be generated since after Wh-Movement the (simplified) structure would be as in (2.29):

2.29) [$_S$I asked [$_S$'#[$_{COMP}$[#which men#]][$_S$#[$_{NP}$there#]
[$_{VP}$#are [##][$_{PP}$#in the room#]]]]]

The auxiliary *are* in (2.29) is followed by a series of two word boundary symbols and, therefore, cannot undergo the Monosyllabic Rule. As a result, the stress on *are* is not reduced and it cannot undergo cliticization.

There are a number of problems with this analysis[14] one of which hinges on the very feature which allows ungrammatical sentences like (2.1aii) to be excluded. If boundary markers are left behind by all movement and deletion rules, then a sentence like (2.30a) with a deep structure as in (2.30b) would incorrectly be blocked from undergoing destressing and subsequent cliticization as can be seen from the surface structure in (2.30c):

2.30) a. John is to leave as soon as possible.
b. [$_S$John is [$_S$'#COMP[$_S$#[$_{NP}$#John#][$_{VP}$#leave. . .
c. [$_S$John is [$_S$'#COMP[$_S$#[$_{NP}$Ø#][$_{VF}$#to#leave. . .
d. John's to leave as soon as possible.

After Equi applies to the lower S there will be a series of four boundary markers, one of which will be removed by the redundant boundary symbol convention discussed above. Since there are three boundary symbols between is and the nearest following word with a stressed vowel the conditions for the Monosyllabic Rule are not met. Since is cannot be destressed it also cannot cliticize, thus the sentence in (2.30d) cannot be produced.

As Postal and Pullum (1978) point out, Selkirk's analysis fails (in precisely the same way) for CC as well. Thus a sentence like (2.31a) would have a surface structure as in (2.31b):

2.31) a. I want to go.
b. [$_{S}$I want[$_{S'}$#COMP[$_{S}$#[$_{NP}\emptyset$#][$_{VP}$#to go#]

Since there are extra boundaries between want and to, destressing and cliticization are incorrectly blocked.

The other types of trace theories proposed thus far have similar problems. Those put forth in Chomsky (1976,1977) assume that movement transformations leave traces in surface structure to mark the position of an element before the rule applied. Postal and Pullum (1978) argue, however, that these theories are incompatible with Chomsky's claim that Wh-Movement is successive cyclic since traces will be overgenerated in COMP position. Thus a sentence like (2.24b) would have the (pre-contraction) surface structure in (2.32):

2.32) [[who do you want[[t] to kiss t]

which is not compatible with CC. Chomsky and Lasnik (1977) attempt to correct this prediction by proposing a rule to freely delete material in COMP positions. However, as Postal and Pullum note, since the COMP node itself is not pruned by this rule want and to are still not structurally adjacent and, therefore, cannot cliticize.

Chomsky (1980) deals with this problem in another way. He argues that traces left by Wh-Movement in non-COMP positions in the clause are case-marked traces and count as syntactic material whereas traces in other positions do not. Since these case-marked traces count as syntactic material, they block contraction. By including this abstract feature, Chomsky is able to distinguish between unbounded dependencies (which do not allow contraction across a t) on the one hand, and Raising and Equi constructions (which do allow this contraction) on the other. Also accounted for is the possibility of cliticization in sentences like (2.32); since the t intervening between want and to is in COMP position it is not a case-marked trace and does not block CC.

There are, however, some problems with these claims. Pullum and Postal (1982) argue that Chomsky's assumptions make it impossible for any dialect of English not to have case marked traces and thus does not account for "liberal" dialects which accept cliticization in sentences in which a marked trace should intervene between the verb and to. In such a dialect sentences like (2.25b) are perfectly fine:

2.25) b. Who do you wanna kiss you?

Furthermore, they point out that since none of the examples in (2.6)-(2.9), repeated below, involve the intervention of a case-marked trace between the want and to, Chomsky's theory fails to account for why cliticization is blocked in each case:

2.6) a. To regret what one does not have seems like to want.
b. ?It seems like to want to regret what one does not have.
c. *It seems like to wanna regret what one does not have.

2.7) a. I don't want anyone [who continues to want] to stop wanting.
b. *I don't want anyone [who continues to wan]na stop wanting.

2.8) a. I want to dance and to sing.
b. *I wanna dance and to sing.

2.9) a. I don't need or want to hear about it.
b. *I don't need or wanna hear about it.

Obviously in these cases mere reference to the position of case marked traces is not enough; one must also take into account other aspects of clause structure, something that Chomsky does not do. Pullum and Postal themselves argue that the underlying failure of trace theories stems from "an unwarranted and unjustified assumption made at the outset and apparently never questioned by TT [trace theory] advocates. This is that linear contiguity is fundamental to the description of contraction" (p. 130). They, however, claim that adjacency is not the primary prerequisite to contraction and propose the following "relational generalization":

2.33) A contraction trigger V can have a contracted form with infinitival to only if:
a. to is the main verb of the initial direct object complement of the matrix clause whose main verb is V;
b. the final subject of the complement is identical to the final subject of the matrix.

If, however, adjacency is not a primary prerequisite we would expect sentences such as (2.34a) to allow contraction since it satisfies both of the conditions specified in (2.33). As we can see from (2.34b), CC is not acceptable here[15]:

2.34) a. I want very much to finish this chapter.
b. *I wanna very much finish this chapter.

From this we must conclude that conventional wisdom is correct after all and linear contiguity is in fact a necessary part of the condition on CC.

2.4 Towards a GPSG Approach.

The treatment of AR and CC I am going to argue for here is, more or less, a trace analysis too, albeit one that refers to clause structure as well. The difference between my analysis and other such analyses is that in a Generalized Phrase Structure grammar different predictions are made about which syntactic structures contain traces or gaps. Thus a GPSG analysis avoids the problem found in transformational treatments of how to distinguish the movement and deletion rules which block cliticization from those that don't. In the next section I will briefly outline the basic tenets of GPSG and show how they can lead to a simple and elegant statement of the conditions governing AR and CC in the dialects discussed here.

3. A GPSG Analysis of the Data

One difficulty in presenting a unified account of AR and CC within GPSG is that the framework itself has been through a number of extensive revisions in a relatively short amount of time. In the following section I will briefly summarize the most recent version of GPSG as presented in Sag and Klein (1982) and Gazdar and Pullum (1982). Throughout this chapter I will attempt to standardize the varying notation as much as possible while maintaining the basic content of the rules; though I adopt the familiar S/NP/VP symbols whenever possible for perspicuity, it should be remembered that GPSG embraces an X-bar philosophy. When necessary I will use the symbol "a" to stand for the Greek letter alpha and "b" for the letter beta.

3.1 The Framework.

GPSG is a surfacy theory of generative grammar in which structural descriptions are assigned to sentences solely on the basis of phrase structure rules; no use is made of transformations or coindexing devices and only one level of structure is defined. The set of immediate dominance (ID) rules are the syntactic basis of a GPS grammar. ID rules have the form:

$$\langle n;\ A \rightarrow B,\ C,\ D\rangle$$

where B, C and D are the categories that A dominates and n is a rule number which acts as a subcategorization feature on any lexical items introduced by the rule. The relative order of B, C and D is given by the set of linear precedence (LP) statements. An example of a LP rule of English would be:

$$NP > PP$$

This says that in any ID rule which introduces both an NP and a PP, the NP will always occur before the PP. In order for a PS rule to be included in the grammar it must be consistent with at least one ID statement and with every LP rule.

Perhaps the most intriguing aspect of GPSG is its use of a metagrammar to capture the generalizations that hold between ID rules and govern their operation but which are not expressed directly with the ID-LP statements. The metagrammar uses two types of devices to capture these generalizations: a set of metarules and a set of rule extension principles. Metarules are a means of expanding the set of ID rules in a rule-governed way; that is, they map ID rules into new ID rules. Metarules have the general form indicated below:

$$3.1)\quad \begin{array}{l} a \rightarrow b_1,\ .\ .\ .\ ,\ b_n \\ \qquad\quad \downarrow \\ a' \rightarrow b'_1,\ .\ .\ .\ ,\ b'_n \end{array}$$

This is interpreted as saying that if the ID rule $a \rightarrow b_1, . . . , b_n$ is in the grammar then the ID rule $a' \rightarrow b'_1, . . . , b_n$ will also be in the grammar. Since by convention rule numbers are preserved under metarule application they are not specifically mentioned in (3.1)[16].

Rule extension principles "flesh out" these schematic ID rules into fully specified PS rules complete with semantic interpretations. These principles are of two types: rule translation principles and feature instantiation principles. The rule translation principles predict the form of semantic translation rules on the basis of the ID rules and the semantic types assigned to the categories they contain. They thus provide a mapping from ID rule doubles, consisting of a rule number and ID rule, into ID rule triples, which contain in addition Montague-like translation formulae.

Features play a very important role in the GPSG framework. In fact, in the most recent versions of the theory much of the work previously done by metarule is now handled by the feature system and the rules which govern feature assignment (i.e. the feature instantiation principles). Not surprisingly the feature system in GPSG has become quite complex[17]. As is also the case with current versions of transformational grammar, syntactic categories in GPSG are not seen as simple unanalyzable node lables but are instead assigned an internal structure consisting of features. The major innovation in the GPSG system is the idea that these features may take other features as their coefficients[18]. Thus the structure of features is defined as follows:

3.2) A feature consists of a feature name optionally followed by one or more features or feature names. Features begin with a left bracket and end with a right bracket. (Gazdar and Pullum (1982), p.3)

Syntactic categories are simply a type of feature, in particular one whose feature name is CAT or CAT'. The internal structure of CAT and CAT' is given below:

3.3) a. [CAT' CAT FOOT]
b. [CAT BAR HEAD]

The feature BAR indicates the phrasal level of the category in an X-bar syntax; it takes as its coefficient a number from 1 to 3 or the feature LEXICAL. For purposes of subcategorization, rule numbers are assigned as the value of the feature LEXICAL. The feature HEAD consists of the syntactic information that is shared between phrases and their heads. This information is divided between the features MAJOR and MINOR as shown in (3.4):

3.4) a. [HEAD MAJOR MINOR]
b. [MAJOR {+N,-N} {+V,-V}]
c. [MINOR AGR CASE . . .]

The feature FOOT contains information about other types of syntactic dependencies that hold between phrases[19]. The internal structure of FOOT and its coefficients is shown in (3.5):

3.5) a. [FOOT SLASH WH REFL]
b. [SLASH CAT]
c. [WH AGR WHMOR]
d. [REFL AGR]

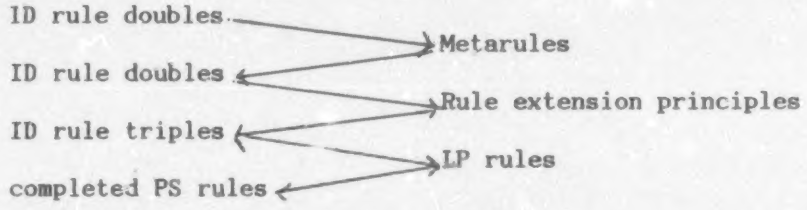
The FOOT feature SLASH is used to encode information about gaps in unbounded dependencies, it takes as its value a category. WH is used in the treatment of Wh expressions; it takes two other features as its coefficients, an agreement feature (AGR) and a feature to encode the morphological type of the Wh word (WHMOR). The feature REFL marks reflexive expressions and also takes AGR as its value. We will have more to say about FOOT features later.

Feature instantiation principles are responsible for ensuring the proper distribution of features in rules. They can be thought of (Sag and Klein (1982), p. 97) as "axioms that must be satisfied by an IDR triple if it is to be an instantiated extension of a given IDR double". Feature values can be assigned in a number of ways: they can be specifically mentioned in an ID rule or metarule, they can be freely assigned in accordance with any default values an item may have (for example, an NP in English is [-CASE], i.e. accusative, unless otherwise specified), or they can be set equal to some other set of features by virtue of special conventions. The special conventions we will be most interested in here are the Head Feature Convention (HFC) and the Foot Feature Principle (FFP).

To put it very simply, the HFC requires the coefficients of HEAD in the mother category and the head daughter to be the same. The "head daughter" is identified on the basis of syntactic category and bar level. For example, given a phrase X'' the head daughter will be either an X'', an X' or an X that it immediately dominates. If X'' dominates more than one of these then the one with the fewest bars will be the head; if it dominates none of these then X'' will have no head[20].

The FFP is responsible for the distribution of FOOT features. Again, very simply put, the FFP says that any FOOT features not assigned to daughters by specific rules must also appear on the mother node. There is nothing to prevent more than one daughter from carrying the same value for a FOOT feature or from carrying different FOOT features altogether, though they are blocked from having different values for the same FOOT feature since there would be no way to encode this on the mother node. Thus, for example, a VP cannot simultaneously have both an NP gap and a PP gap since SLASH can have only one value for CAT.

Rule translation and feature instantiation are two aspects of the mapping from ID rule doubles to ID rule triples. Sag and Klein (p. 98) point out that since both can affect how constituents are linearized in a language the set of rule extension principles must operate before the LP statements. Their view of how the grammar is organized is given below:



3.2 Auxiliaries in GPSG.

My approach to auxiliaries is basically the same as that presented in Gazdar, Pullum and Sag (1981) with one small exception having to do with the treatment of the copula. In that work a verb that is [COP, AUX] can take any of the following complements:

3.6) VP[PRP] is going
VP[PAS] is given
VP[INF] is to leave
VP[PRD]

where a VP[PRD] "merely consists of a predicational X" [i.e. XP]" (GPS, p. 9). I will simplify this somewhat and say that any [COP] verb that takes an XP[PRD] as its complement is also an [AUX], where an XP[PRD] can be any of the following: AP[PRD], VP[PRD, PAS], VP[PRD, PRP], VP[PRD, INF], NP[PRD], PP[PRD]. This results in slightly different tree structures (i.e. no dominating VP node for AP's, NP's, and PP's) and is more in keeping with more recent GPSG works. Notice that by this definition the verb in (3.7a) is not an auxiliary, since its complement is an S (i.e. V''') rather than an XP, and therefore does not undergo AR, as we can see from (3.7b):

3.7) a. The fact is Pita left.
b. *The fact's Pita left.

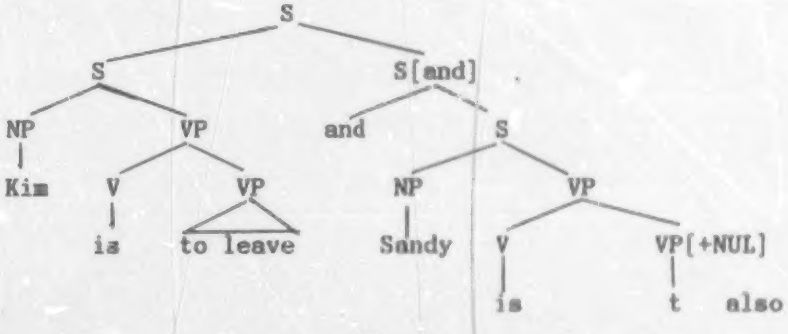
Thus, sentences like (3.7b) will not be considered in our later discussion of the conditions under which AR takes place. Again, where necessary I will modify rule notation to be consistent with this treatment of the copula and its complements.

3.3 The Distribution of Traces in GPSG.

As we saw in chapter 2, the problem with transformational analyses of AR and CC is that they fail to distinguish in a general way between operations that block contraction-- such as Topicalization, VP Deletion and most forms of Wh-Movement--and those that don't--i.e. Equi and There Insertion. A GPSG analysis of the same data does not run into this problem because in GPSG there is a natural distinction between the two sets of constructions: the GPSG equivalents of the former involve the introduction a phonologically null element while the GPSG equivalents of the latter do not. Thus the distribution of these null elements can be used to state the conditions governing the application of AR and CC.

Since GPSG is a non-transformational monostratal theory, null elements do not arise through the operation of movement or deletion rules. Instead, the distribution of traces is governed by the feature system and metarules. Categories which are marked with the feature [+NUL] do not receive a phonological representation and are, therefore, trace elements. Thus, the GPSG equivalent of VP Deletion is achieved simply by assigning a VP this feature. Thus a sentence like (2.1dii) would have the pre-reduction structure in (3.8):

3.8)



where t is an abbreviation for VP[+NUL][21]. Since trace elements retain their other category features, null categories will be linearized by the LP statemants just like their non-null counterparts.

In my analysis, traces are also introduced by a version of Slash Termination Metarule 1 (STM1), one of the rules used in Gazdar, Klein, Pullum and Sag (1982) (GKPS henceforth) to "eliminate" unbounded dependencies. This version of STM1 is given below[22]:

3.9) STM 1
a -> W, b[-CASE]
↓
a/b -> W, t

(3.9) says that given an a consisting of anything at all (i.e. W) and a b that is [-CASE], there exists in the grammar a rule that allows an a that has b as its coefficient for SLASH to dominate a W and a trace. The "a/b" notation used here is simply shorthand for the actual feature specification of the mother node which would be [CAT' a [FOOT [SLASH b]]].

Following the analysis in GKPS, the rules responsible for introducing unbounded dependencies are contained in the set of ID rules. Two such rules are given in (3.10):

3.10) a. S -> a, S/a
b. S -> PP, VP[there]/PP

By itself, the rule in (3.10a) is responsible for topicalization constructions such as:

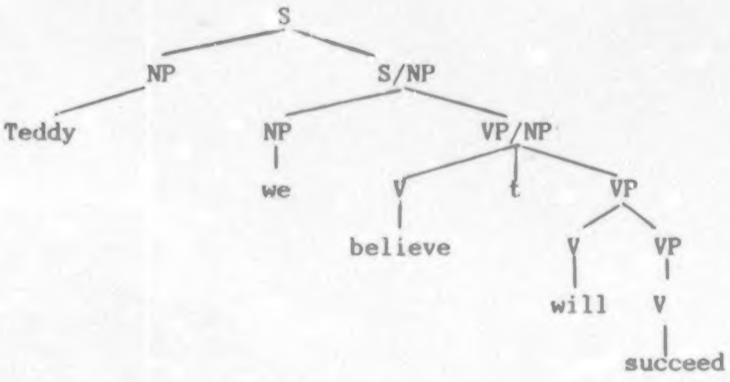
3.11) Teddy, we believe will succeed.

In conjunction with other ID rules and The FOOT feature WH it also accounts for most of the effects of Wh-Movement in a transformational analysis. The rule in (3.10b) is responsible for sentences like (3.12). The feature [there] indicates that the VP is the kind that could take an existential subject as in (3.13):

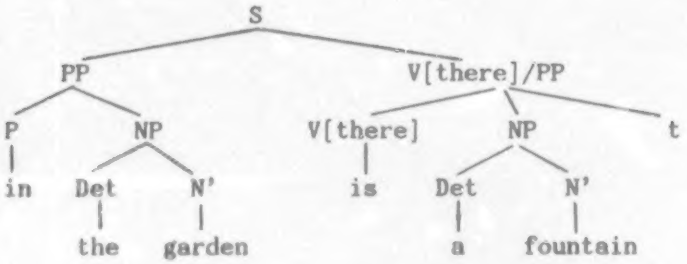
3.12) In the garden is a fountain.
3.13) There is a fountain in the garden.

Since the FOOT feature SLASH takes as its value the category of the "missing" element in an unbounded dependency, this information will be carried through the tree from the point of introduction to the point of elimination by the FFP. Thus given STM1, the rules in (3.10) and the FFP, sentences like (3.11) and (3.12) will be assigned the following structures:

3.14)



3.15)



Since, however, GKPS restrict metarules so as to operate only on ID rules that introduce lexical items, STM1 could not apply in the production of a sentence like:

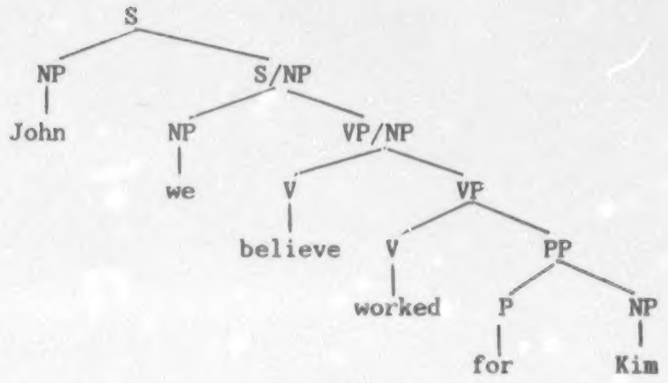
3.16) John, we believe worked for Kim.

because the rule STM1 would have to apply to would be the rule expanding the S complement of believe as an NP and a VP. This application is blocked since neither NP nor VP is a lexical category. To account for sentences like this, among others, GKPS propose a second STM rule. This rule replaces and generally supersedes the one given in Gazdar (1981) which allowed sentential categories that were missing an NP to be replaced by a VP. Like this rule, STM2 does not involve the introduction of a trace element, rather it allows the remnants of an embedded clause to be "liberated" into a higher clause. This second slash termination metarule is given in (3.17):

3.17) STM2
a -> W, b
↓
a/I -> W, J   where "b -> I, J" is a nonlexical rule

STM2 says that if the grammar has an ID rule which introduced a b and b can dominate I and J, where neither I nor J is a lexical category, then the grammar also has a rule in which b is replaced by J and I is assigned as the value of the mother node's SLASH feature. Given this rule, the sentence in (3.16) would be assigned the structure:

3.18)



Given this account of the distribution of traces in GPSG, statements of the conditions governing the application of AR and CC fall out directly. In the next two sections, I will show how an analysis of the data discussed in chapter two can be devised using a GPSG syntax as a base. Special emphasis will be placed on accounting for dialect variation in a simple and natural way without any ad hoc devices.

3.4 AR in GPSG.

Given a GPSG syntactic framework, the condition on AR in the most liberal dialect is quite easy to state: auxiliaries can contract if they are immediately followed by phonologically non-null material from their own constituents. More restrictive dialects, such as Kaisse's, require an additional condition on possible host elements as well. These dialects will be discussed further below. Notice, however, that I do not attempt to give here a formal statement of what the AR rule looks like. This is because, as we saw in chapter one, both AR and CC are not and can not be syntactic rules themselves. We will return to this issue and how it is predicted by a GPSG framework in section 3.6. Notice also that since cliticization rules are not located in the syntactic component of the grammar they need not be subject to the same types of restrictions as syntactic rules. Just what the general restrictions on cliticization rules are is a topic for future research.

Given this constraint, sentences like (3.19a) will be prevented from undergoing reduction regardless of whether the verb is analyzed as an auxiliary:

3.19) a. I think therefore I am.
b. *I think therefore I'm.

The sentence in (3.19b) is ungrammatical because nothing, not even a trace, follows the auxiliary in its constituent. Even if the verb in (3.19a) is not an auxiliary (which it no doubt isn't), this wording is required on independent grounds to account for sentences like (2.11):

2.11) a. He is, I should think, a bit tired.
b. *He's, I should think, a bit tired.

In this case the verb is an auxiliary by our definition since it takes an XP[PRD] complement (namely an NP) but it still doesn't allow reduction. This is because the material that immediately follows it is not contained in its clause, rather it is a separate S which is inserted parenthetically. Thus, (2.11b) is correctly excluded by our statement of the conditions allowing AR.

Unlike Lakoff's analysis, the reduction on AR given above does not run into difficulties with sentences like (2.2):

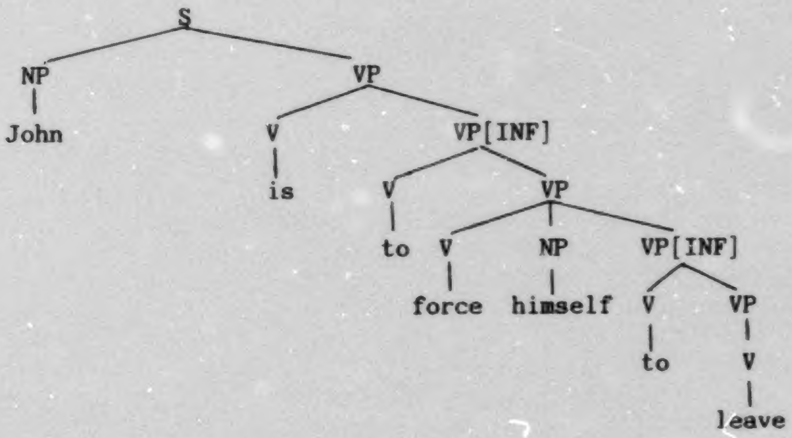
2.2) John's to force himself to leave.

simply because GPSG makes different predictions than transformational grammars about how such sentences are produced. While Lakoff's framework entailed the application of Equi NP Deletion and to Insertion to transform an embedded S into an infinitival complement, verbs in GPSG subcategorize for infinitival complements directly via rules like (3.20):

3.20) VP -> V VP[INF]

Thus the sentence in (2.2) would be assigned the pre-reduction structure in (3.21):

3.21)



Since phonologically non-null material immediately follows the be in its constituent cliticization is possible.

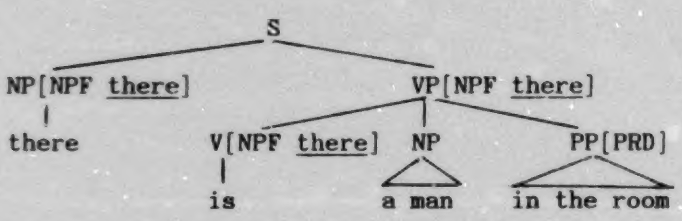
Similarly, under the analysis given in Sag and Klein (1982), so-called There Insertion sentences involve the interaction of subcategorization and agreement rather than string manipulation. Their analysis relies on rules like the following, adapted from the original X-bar notation (Sag and Klein, p. 103):

3.22) < 7; NP[NPF a] -> a >, where a ∈ {it, there}
3.23) < 12; VP -> V[-PRP, NPF there], NP, XP[PRD] >, where
V[12] = {be}

NPF is a type of agreement feature which insures that dummy NP's appear in the appropriate structures. Since agreement is stated between subject NP's and their VP's[23] and subsequently carried through the tree by the HFC, these rules result in pre-reduction structures such as (3.24) for sentences like (2.1ai):

2.1) a. i. There's a man in the room.

3.24)

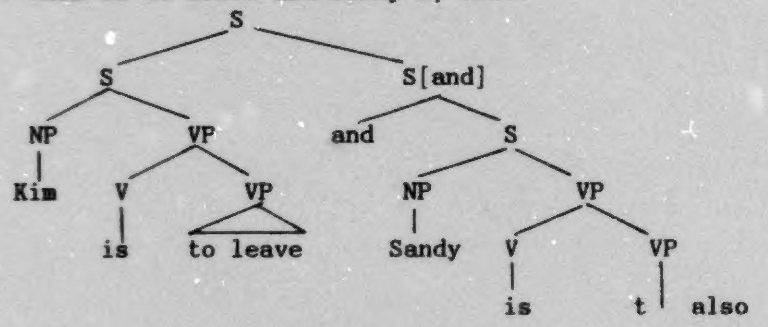

S
NP[NPF there]
VP[NPF there]
there
V[NPF there]
NP
PP[PRD]
is
a man
in the room

Unlike some transformational analyses, there is no disturbance in the material following the auxiliary and thus no need, as with Kaisse (1983a), to refer to more than one level of structure in order to account for the ability of AR to apply here.

While Equi and existential there sentences do not involve phonologically null elements, the GPSG equivalent of VP Deletion, as we have seen, does. Therefore, if a VP that is assigned the feature [+NUL] immediately follows an auxiliary that auxiliary cannot undergo AR. Thus a sentence like (2.1di) will not be grammatical since it is assigned the pre-reduction structure given in (3.8):

2.1) d. i. *Kim is to leave and Sandy's, also.

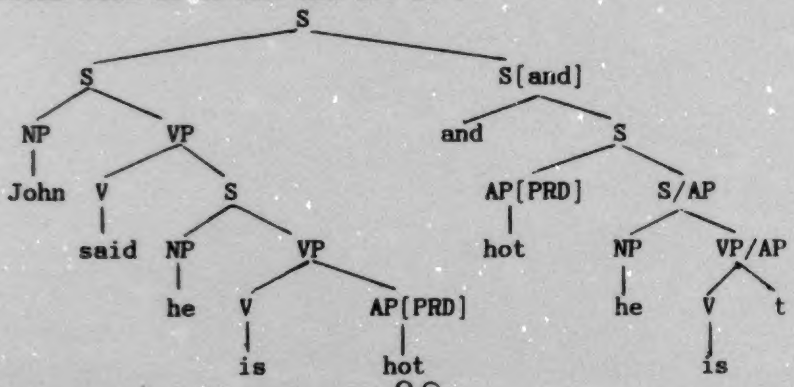
3.8)


S
S
S[and]
NP
VP
and
S
Kim
V
VP
NP
VP
is
to leave
Sandy
V
VP
is
t
also

AR can also be blocked from applying in topicalized sentences since STM1 will be used to eliminate the unbounded dependency. Thus a sentence like (3.25a) could not undergo reduction in the second conjunct since it would be assigned the structure in (3.26), with a phonologically null element (a trace) after the auxiliary:

3.25) a. John said he is hot and hot he is!
b. *John said he is hot and hot he's!

3.26)


S
S
S[and]
NP
VP
and
S
John
V
S
AP[PRD]
S/AP
said
NP
VP
hot
NP
VP/AP
he
V
AP[PRD]
he
V
t
is
hot
is

STM2, which does not introduce traces, could not be used to eliminate the dependency here since the rules expanding adjective phrases introduce lexical items (cf (3.17)).

The rules responsible for topicalizations also interact with the FOOT feature WH and our constraint in such a way as to explain the ungrammaticality of sentences like (2.1aii) and (3.27):

2.1) a. ii. *I asked which men there're in the room.
3.27)*The restaurant in whose cellar that wine's will be the most popular.

In GKPS, the ID rules which introduce embedded questions and relative clauses are the following:

3.28) VP -> V[18], S[+Q]
3.29) NP -> NP, S[+R]

where Q is the value WHMOR takes for interrogatives and R the value it takes for relatives. [+Q] and [+R] are used as abbreviations for the features [WH AGR [WHMOR Q]] and [WH AGR [WHMOR R]] respectively. The rules expanding S[+R] and S[+Q] are the result of instantiating independently needed S expansions with these features by the FFP[24]. Thus, since we have the rule in (3.30a) we will also have the rules in (3.30b):

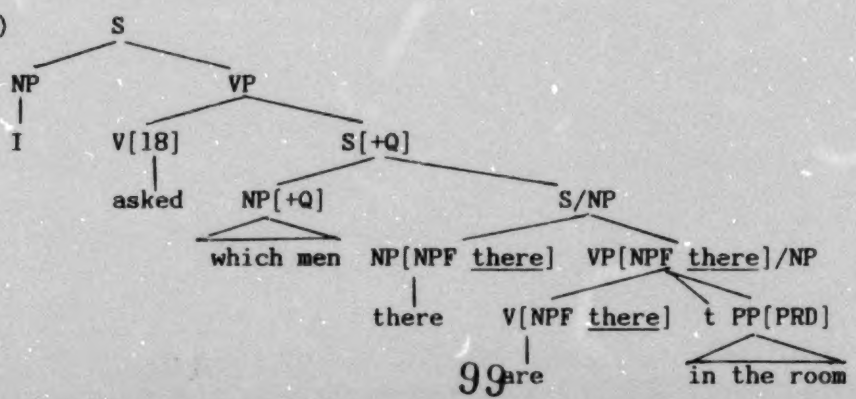
3.30) a. S -> NP, VP
b. i. S[+Q] -> NP[+Q], VP
ii. S[+R] -> NP[+R], VP

Similarly, (3.31a) and (3.32a) will legitimate rules like (3.31b) and (3.32b):

3.31) a. S -> PP, VP[there]/PP
b. i. S[+Q] -> PP[+Q], VP[there]/PP
ii. S[+R] -> PP[+R], VP[there]/PP
3.32) a. S -> a, S/a
b. i. S[+Q] -> a[+Q], S/a
ii. S[+R] -> a[+R], S/a

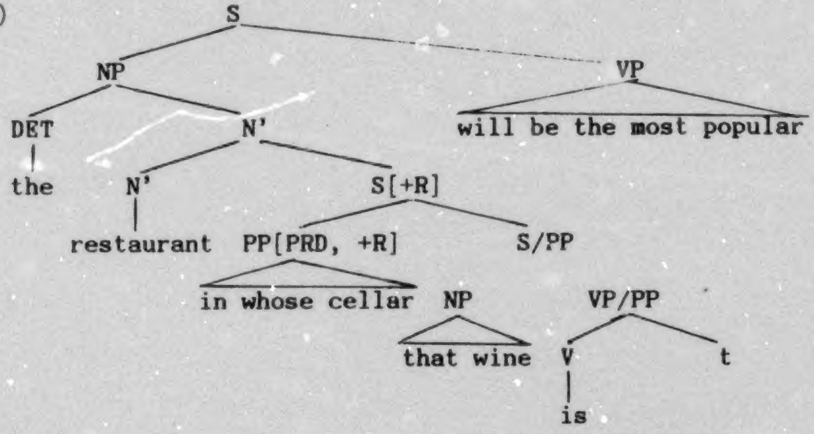
Given the rule in (3.32bi), the pre-reduction structure for (2.1aii) would be as follows, where a = NP:

3.33)



Since the unbounded dependency introduced in connection with the NP which men is eliminated by STM1 a trace is left. Since this trace has all the category features of an NP, it will be linearized into the position following the verb are by the LP statements, thus preventing AR from applying. Sentences such as (3.27) are blocked for similar reasons, as we can see from the structure in (3.34):

3.34)

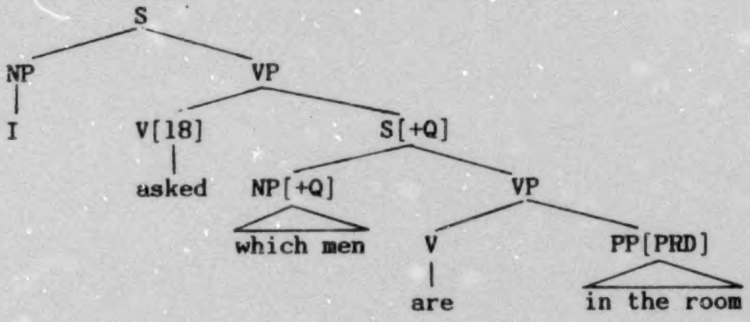

S
NP
VP
DET
N'
will be the most popular
the
N'
S[+R]
restaurant
PP[PRD, +R]
S/PP
in whose cellar
NP
VP/PP
that wine
V
t
is

The relative clause expansion used here is the one given in (3.32bii) with a = PP.

Notice, however, that superficially similar sentences such as (3.35a) and (3.36a) will allow AR to apply since their embedded sentences are expanded by the rules in (3.30b), as shown in (3.35b) and (3.36b):

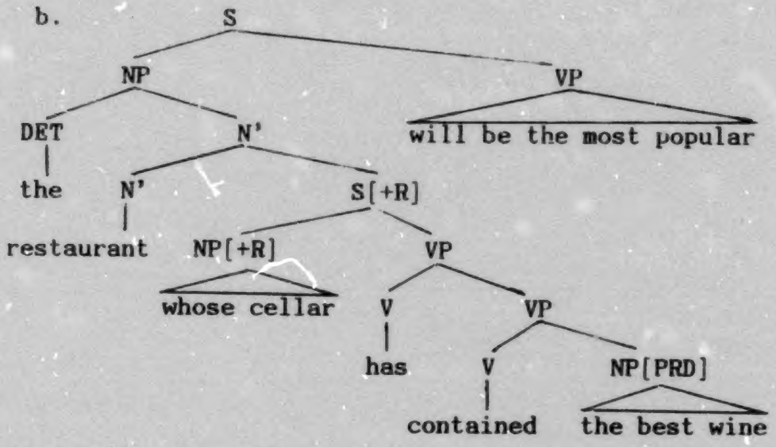
3.35) a. I asked which men are in the room.

b.


S
NP
VP
I
V[18]
S[+Q]
asked
NP[+Q]
VP
which men
V
PP[PRD]
are
in the room

3.36) a. The restaurant whose cellar has contained the best wine will be the most popular.

b.



Since no unbounded dependency is introduced (and VP Deletion is not involved) no categories are marked [+NUL] and the verb is followed in its own clause by a phonologically non-null element. Therefore, AR is possible.

Unlike the earlier analysis of unbounded dependencies presented in Gazdar (1981), however, the one given in GKPS _only_ addresses the issues of embedded sentences and relative clauses; no mention is made of root _wh_ questions such as (3.37)-(3.39):

3.37) Who is Pita?
3.38) In which garden is a statue of Pita?
3.39) In which park did Pita say the concert is?

In fact, given the assumptions in GKPS it is difficult to see how they _could_ account for such sentences in a reasonable and regular way. One problem is that the interrogative feature [+Q] does not distinguish between complement questions and non-complement questions and their concomitant word order differences. We could add a feature [± C] (i.e complement) to capture this distinction and the feature co-occurrence restriction:

$$[\underline{a}\ \mathrm{C}] \supset [-\underline{a}\ \mathrm{INV}]$$

to insure that [-C] questions were inverted and [+C] question were [-INV], i.e. (3.40) but not (3.41):

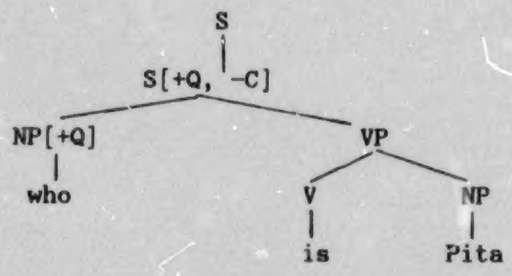
3.40) a. Who can Pita see?
b. I know who Pita will see.
3.41) a. *Who Pita can see?
b. *I know who will Pita see.

but even this would not give us a completely adequate account of root _wh_ questions. This is because the FOOT feature [+Q] works the same way as the FOOT feature SLASH (GKPS, p. 54) and, as such, must be specifically introduced by an ID rule. Thus, if we wish to take advantage of the prediction of feature instantiation, as we did with relative clauses and embedded questions, we would have to propose a rule like the following:

3.42) S -> S[+Q, -C]

Notice, however, that this rule produces some very peculiar tree structures:

3.43)



There is no independent justification for the extra S node dominating the S[+Q, INV], rather its only reason for existence is to allow us to introduce the [+Q] feature.

We could, of course, simply list each of the rules expanding an S[+Q, -C] separately in the grammar as shown below:

3.44) a. S[+Q, -C] -> NP[+Q], VP
b. S[+Q, -C] -> PP[+Q], VP[there]/PP
c. S[+Q, -C] -> a, S[INV]/a
etc.

but this would result in a great deal of redundancy in the ID rules and fails to capture generalizations about the feature system and the structure of root wh sentences. It would be better if the grammar somehow predicted the existence of the rules in (3.44).

A possible solution to these problems would be to give up GKPS's stipulation that metarules may only apply to lexical ID rules and introduce root wh question expansions via the following[25]:

3.45) S[+Q] -> H, W
↓
S -> H[INV], W

where the default value for WH is assumed to be null. This metarule says that if the grammar has a rule that expands an S[+Q] as a head and its complements, then the grammar will also have a rule that expands a regular S in the same way except that its head will be marked with the inversion feature[26]. Thus, since the grammar will have the rules expanding S[+Q] given in (3.30bi), (3.31bi) and (3.32bi) the grammar will also have the rules in (3.46)[27]:

3.46) a. S -> NP[+Q], VP[INV]
b. S -> PP[+Q], VP[there, INV]/PP
c. S -> a, S[INV]/a

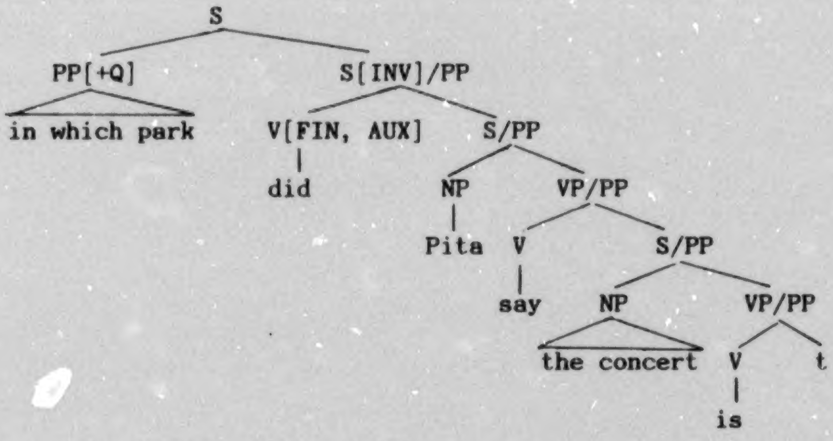
Recall that the notation "a/b" is simply an abbreviation for [CAT' a [FOOT [SLASH b]]] and that head daughters are chosen on the basis of bar level and syntactic category. Since having a value for SLASH in no way affects the category or bar level features of a node, an a/b can qualify as the head of a phrase as long as the other criteria are met. Thus the VP in (3.46b) and the S in (3.46c) will be assigned the inversion feature by (3.45) despite the fact that they are slashed categories (i.e. a slashed VP is still a VP, etc.).

The rule in (3.46c) will interact with the output of the Subject Auxiliary Inversion (SAI) metarule, shown below, to produce sentences like (3.39)[28]:

3.47) SAI Metarule
VP [FIN, AUX] -> V, VP[PRD, a]
↓
S[INV] -> V[FIN, AUX], S[a]

The SAI metarule says that if the grammar contains an ID rule expanding a finite verb phrase as an auxiliary verb and its VP complement, then the grammar also has a rule expanding an S with the inversion feature as a finite auxiliary verb and an S with the same subcategorization features as the VP complement of the auxiliary in the input rule. Thus, the structure for a sentence like (3.39) would be as in (3.48), where a in (3.46c) takes the value PP:

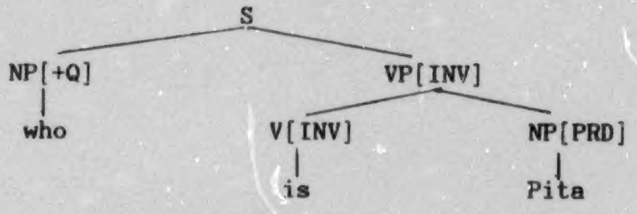
3.48)



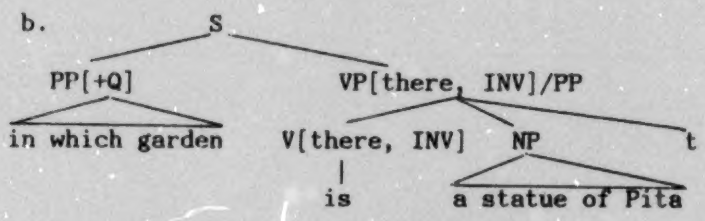
Since we are assuming that unless otherwise specified no value for the interrogative feature is assigned, we do not need to worry about introducing the nodes which expand as root wh questions in other ID rules, as we did with embedded questions. Also, if we assume that the default value for the feature [INV] is negative, we can insure that embedded questions like (3.41b) can not be generated (alternately we could specify the value [-INV] in the rule that introduces embedded questions itself, i.e. (3.28)).

As shown above, the rule in (3.45) also assigns the [INV] feature to the VP's in (3.46a,b). By the HFC this feature will be passed on to the V's these VP's dominate to produce structures like (3.49):

3.49) a.



b.



This would be fine except for the feature co-occurrence restriction proposed in Gazdar, Pullum and Sag (1981) shown below:

3.50) $[\text{INV}] \supset [\text{AUX, FIN}]$

(3.50) says that if something has the feature [INV], then it will also be an auxiliary and be finite. Given this restriction we would not be able to produce a sentence like (3.51):

3.51) Who loves Pita?

since *loves* is not an auxiliary. This FCR was proposed to prevent sentences like (3.52) from being legitimized by the SAI metarule:

3.52) *Loves Pita to sing?

If, however, we formulate the SAI metarule as in (3.47), with the features [AUX] and [FIN] specifically stated on the verb, sentences like (3.52) will be blocked and the FCR in (3.50) made superfluous. Thus we can dispense with (3.50) without making false predictions[29].

Nor will allowing the feature [INV] to sometimes appear on V's that don't begin a sentence interfere with the treatment of morphological irregularities such as the following:

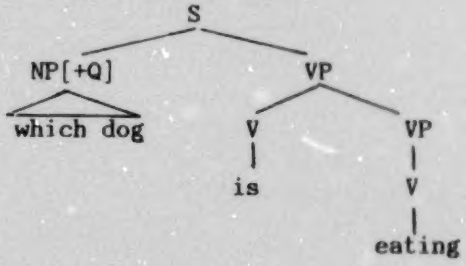
3.53) a. *I amn't going.
b. *Amn't I going?
c. *I aren't going.
d. Aren't I going?

As Gazdar, Pullum and Sag observe (p. 31), this paradigm can be accounted for simply by stipulating in the lexicon that the first person singular present tense copula has no [-INV, +n't] form and that its [+INV, +n't] form is *aren't*. Since the only time that V's that don't begin a sentence are marked as being inverted is in connection with *wh* words or phrases, and since *wh* words and phrases are always *third* person never first person, this lexical restriction will not be affected and we will not incorrectly predict that sentences like (3.53c) are grammatical.

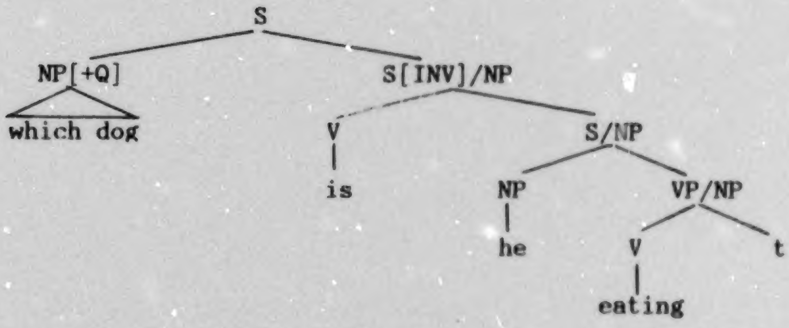
This analysis of root wh questions is also consistent with the facts surrounding the application of AR. Thus, in the liberal dialect, the sentences in (3.37) and (3.38) will allow reduction since in each case the auxiliary verb is immediately followed in its constituent by a phonologically non-null element, as we can see from the structures in (3.49a) and (3.49b) respectively. A sentence like (3.39), however, will not permit AR since the auxiliary precedes a trace element, as shown in (3.48). Similarly, both the sentences in (3.54) will allow AR since they are assigned the corresponding structures in (3.55):

3.54) a. Which dog is eating?
b. Which dog is he eating?

3.55) a.



b.



The importance of the difference between these two structures will become apparent in the following section.

3.4.1 Conservative Dialects.

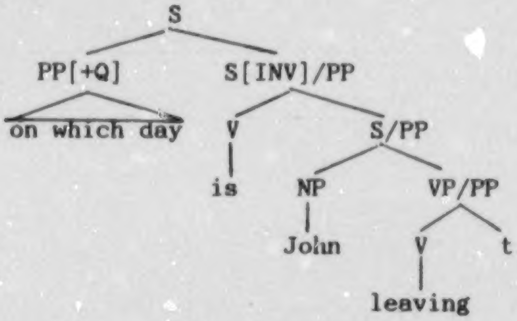
While the constraint on AR given at the beginning of section 3.3 correctly predicts the facts about the dialect described there, other dialects are not quite as "liberal" with regard to where they permit reduction to occur. The dialect described in Kaisse (1981), for example, differs from more liberal dialects in that it does not allow reduction in sentences such as the following, repeated here from section two:

3.56) Not only's Louis smart, he's also a varsity rower.
3.57) On which day's John leaving?
3.58) a. Speaking tonight's a famous reporter.
b. Speaking tonight's been a famous reporter.
3.59) Under this slab's buried Joan of Arc.

Since such sentences are perfectly acceptable in the more liberal dialect, their ungrammaticality here cannot be the result of phonologically null elements being positioned after the auxiliary. There must, therefore, be some other type of constraint at work in these cases. In order to determine what this constraint would be, let us briefly consider the structures such sentences could be assigned in GPSG and what they have in common.

Sentences similar to (3.57) have already been discussed in connection with root wh questions; given the rule in (3.46c), (3.57) would be assigned a structure very much like the one in (3.55b) above:

3.60)



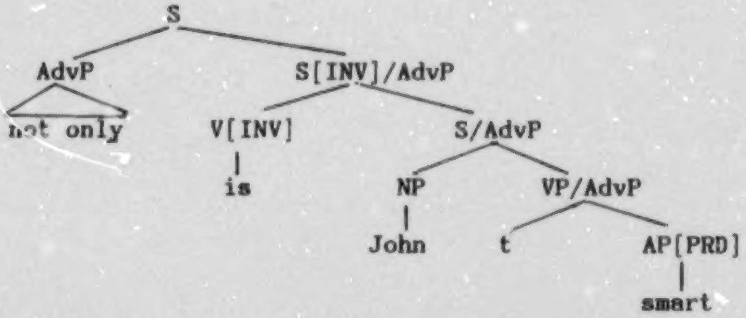
The rest of the sentences listed above, however, have not (to my knowledge) been previously addressed within the GPSG framework.

The analysis of sentences like (3.56) in GPSG, as we shall see, is fairly straightforward. Kaisse, following Emonds (1976), analyses such sentences as being derived from the deep structure in (3.61) by Negative Preposing and SAI:

3.61) Louis is not only smart, he's also a varsity rower.

Assuming the basic correctness of this choice of positioning for the adverb phrase, the corresponding GPSG structure would be as follows:

3.62)



where the "preposing" is achieved by the familiar rules for topicalization. The inverted word order can be guaranteed by means of a FCR on slash introductions such as the following[30]:

[SLASH $\underline{a}$[+Neg]] ⊃ [INV]

Notice that the auxiliary in this sentence does indeed meet the conditions on AR outlined in the preceding section; thus, its ability to reduce in the liberal dialect is explained.

The analyses of sentences like (3.58) and (3.59), however, are a bit more involved. On a transformational analysis (3.58a) would be assigned the deep structure given in (3.63):

3.63) A famous reporter is speaking tonight.

(3.58a) would then be derived by applying a preposing rule to the participal phrase thus triggering a type of inversion. This inversion differs from SAI, which is responsible for sentences like (3.56) and (3.57), in that it applies to what remains of the predicate after preposing rather than just the first auxiliary verb, as we can see from (3.58b). Notice, however, that not all sentences with preposed participial phrases allow this inversion:

3.64) a. Mary saw the mayor holding his breath and counting his ballots.
b. Holding his breath and counting his ballots Mary saw the mayor.
c. *Holding his breath and counting his ballots saw Mary the mayor.

Nor are participles the only preposed phrases which do trigger it:

3.65) a. Snowflakes of all shapes and sizes had fallen from the sky.
b. From the sky had fallen snowflakes of all shapes and sizes.

Since, in transformational terms, preposing the complement of some verbs can result in this type of inversion while preposing the complement of other verbs does not, a metarule to produce such structures in GPSG would need to refer to the verb which subcategorizes for the topicalized phrase in order to determine whether the metarule is applicable. The problem is that this verb need not be the matrix verb, it can be preceded structurally by (other) auxiliaries. Since the inverted order, when allowed, involves the subject and the remnants of its verbal complement, such a metarule would 1) need to refer to a varying number of levels and 2) need to refer to more than two levels of structure. Metarules in GPSG, however, are not permitted to do either of these. Furthermore, if this rule could be written a sentences like (3.58a) would have a trace immediately following the be and thus would be predicted not to undergo AR in the liberal dialect; this is, as we have seen, not a correct prediction. There is, however, an alternative to the metarule approach which not only accounts for the "inverted" word order, it also makes the correct predictions about the positioning of traces.

First of all, notice that verbs like be and fall (but not saw) have something in common other than the fact that when their complements are topicalized this inverted order is found: both be and fall allow existential subjects. Thus, alongside sentences like (3.56) and (3.65) we find sentences like:

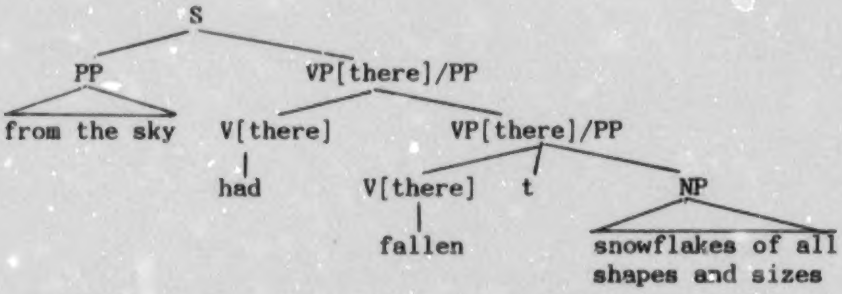
3.66) There is a famous reporter speaking tonight.
3.67) There had fallen from the sky snowflakes of all shapes and sizes.

Since the matrix VP's in these sentences carry the feature [NPF there] in agreement with their subjects, (3.67) can have unbounded dependencies introduced by the ID rule in (3.10):

3.10) S -> PP, P[there]/PP

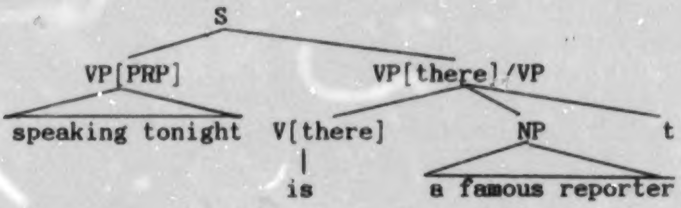
Thus, the sentence in (3.65b) could be assigned the structure:

3.68)

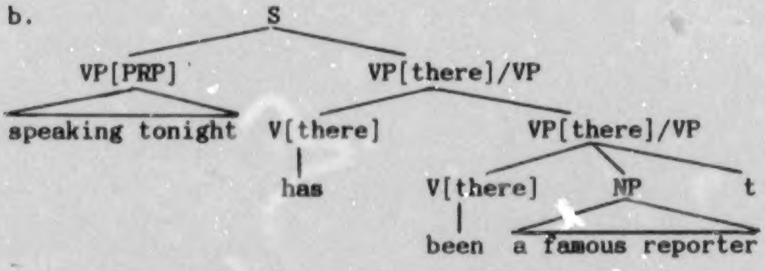

S
PP
VP[there]/PP
from the sky
V[there]
VP[there]/PP
had
V[there]
t
NP
fallen
snowflakes of all shapes and sizes

Sentences like (3.58) can be accounted for by generalizing the rule in (3.10) so as to apply to VP[PRD] as well, yielding structures such as:

3.69) a.


S
VP[PRP]
VP[there]/VP
speaking tonight
V[there]
NP
t
is
a famous reporter

b.


S
VP[PRP]
VP[there]/VP
speaking tonight
V[there]
VP[there]/VP
has
V[there]
NP
t
been
a famous reporter

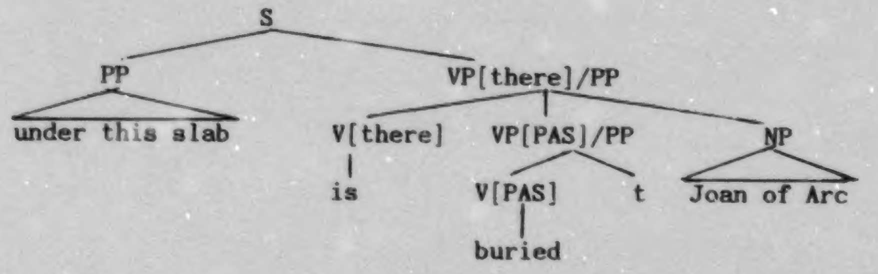
Notice that in each case the initial auxiliary verb is followed in its constituent by a phonologically non-null element. Thus this analysis, unlike the transformational one outlined above, will allow AR to apply here in liberal dialects.

A similar analysis can be proposed for sentences like (3.59). Since there will be ID rules in the grammar licensing the occurrence of sentences like:

3.70) There is buried under this slab Joan of Arc.

then, given the rule in (3.10b), sentences like (3.59) will also be admitted. A likely structure for such a sentence is given in (3.71)[31]:

3.71)



Again, reduction is predicted and found in liberal dialects.

The question then is: given the analyses and judgments presented above, how can the constraint on AR found in this conservative dialect be stated? Since all of the AR sentences that are bad in the liberal dialect are also bad in the conservative dialect, the condition on material following the auxiliary will be needed in both. In order to account for the sentences presented in this section, however, the conservative dialect will also need a means of restricting the set of elements that can serve as host to the cliticization. The constraint proposed by Kaisse (1983b), recall, restricted the host to NP's which c-command the auxiliary. Due to the different structures assigned by our frameworks, her restriction cannot be carried over into this work since it would fail to distinguish between sentences like (3.54a) and (3.54b). As we can see from the structures in (3.55), in both sentences the auxiliary is c-commanded by the potential host and followed by a phonologically non-null element, yet in Kaisse's dialect only (3.54a) allows AR. The correct predictions are made, however, if we instead require the NP host to be commanded by the auxiliary. This in effect prevents an auxiliary in this dialect from attaching outside of its own S and, in addition to blocking AR in sentences like (3.54b), ties in very nicely with Kaisse's later observations about reduced auxiliaries and 2P clitics (Kaisse (1983b)).

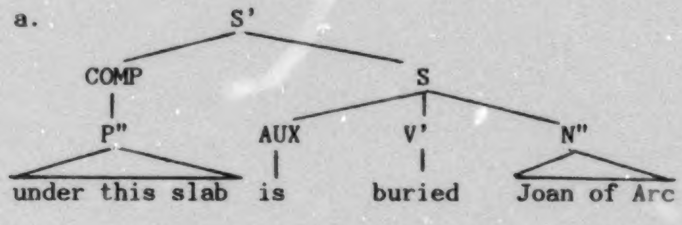
This analysis is superior to either of the ones proposed by Kaisse in that it not only accounts for the relevant grammaticality judgments in both liberal and conservative dialects, but does so with a related set of rules rather than entirely separate ones. As a result the underlying similarity of the two dialects is highlighted. Furthermore, all of this is done without reference to global rules, multiple levels of structure or transformations—extremely powerful devices which are simply not needed in a GPSG syntax. Finally, this approach has the added benefit of giving insights into judgments in the conservative dialect that Kaisse herself could not explain.

Kaisse (1983a) notes (p. 109) that some speakers do not find inversion sentences with preposed PP's such as (3.59) as bad as inversion sentences with preposed participles (cf (3.58)), a fact that she attributes to the relative NP-ness of the two types of phrases. If, however, the greater degree of acceptability found with sentences like (3.59) is a result of the NP-ness of the PP host, we would also expect sentences like (3.57) to be relatively more acceptable as well, since they also have PP hosts. Such is apparently not the case in these dialects. Nor can Kaisse rely on the difference in following context to differentiate the two sentence types since sentences like (3.72) receive similar judgments (Kaisse (1983a), p. 109):

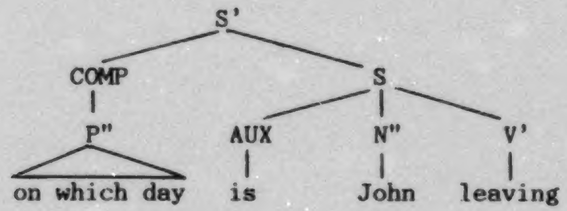
3.72) ?Implicit in your statement's the idea that men are inferior.

A GPSG syntax, however, provides a straighforward explanation of this difference in host status: the PP in (3.57), as we can see from (3.60), fails the condition on hosts on two counts 1) it is not an NP and 2) it is not commanded by the auxiliary seeking to reduce. The PP in (3.59), on the other hand, is commanded by the reducing auxiliary (cf (3.71)). Thus if the NP host condition is relaxed for the speakers in question so as to include PP's, there would be nothing to prevent the auxiliary in (3.59) (or (3.72)) from reducing while AR in (3.57) would still be blocked on structural grounds. No such structural distinction between PP hosts is possible in Kaisse's analysis, however, since in her theory all preposed material is inserted into COMP:

3.73) a.



b.

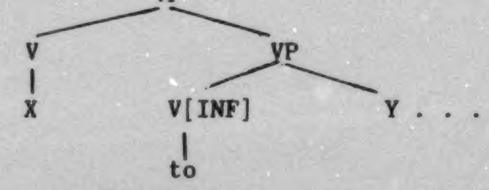


Since the structures in (3.73) are completely analogous (except for the different status of the material following AUX which, as we've seen, cannot be the cause of the grammaticality distinction) no competing explanation of these facts is available.

3.5 CC in GPSG.

As was the case with AR, the syntactic conditions governing the application of CC are relatively easy to state assuming we use a GPSG syntax. Within the majority dialect this condition is as follows: contraction is possible only if the node introducing the trigger verb[32] is the aunt of the node introducing to and they are linearly adjacent. Put more simply, contraction is only possible in structures such as the following[33]:

3.74)



Such a definition entails that in order for CC to take place the triggering verb must c-command the to. Thus, sentences such as (2.6c) and (2.7b), repeated below, are ungrammatical since in these cases want does not c-command to:

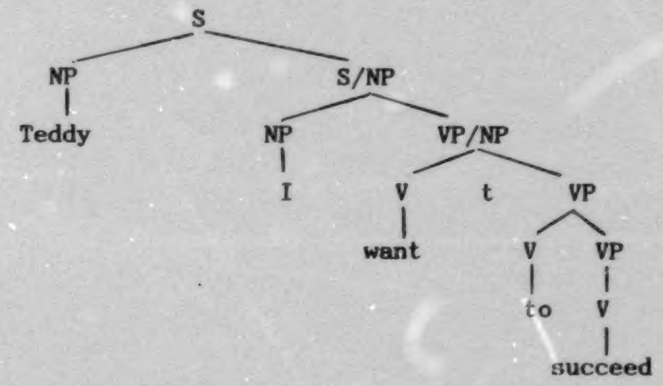
2.6) *It seems like to wanna regret what one does not have.
2.7) *I don't want anyone who continues to wanna stop wanting.

This condition also provides an account of the ambiguity constrasts found between pairs like the following:

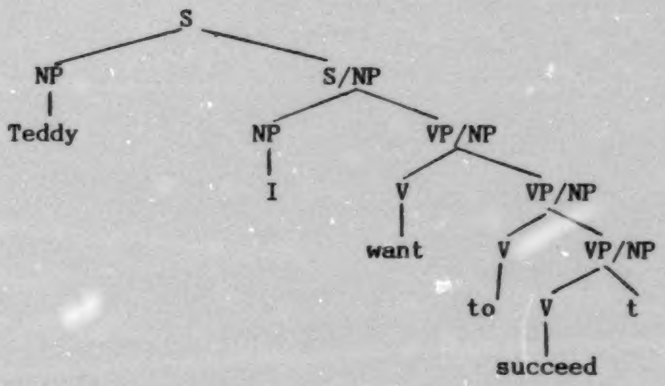
3.75) Teddy, I want to succeed.
3.76) Teddy, I wanna succeed.

The sentence in (3.75) could involve either topicalization from object position in the main clause or topicalization from object position in the embedded clause; since (3.75) has two possible structures it also has two possible interpretations. The sentence in (3.76), on the other hand, has only one interpretation since only one of the structures assigned to (3.75) satisfies the conditions on CC. The structures assigned to (3.75) are shown below:

3.77) a.



b.

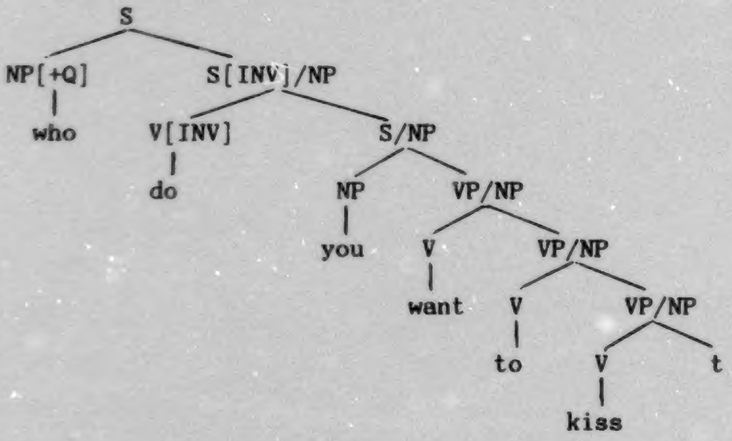


Though the trigger want is the aunt of to in (3.77a) CC is not possible since want does not immediately precede to, there is a trace intervening. In (3.77b), however, there is no such intervening trace and

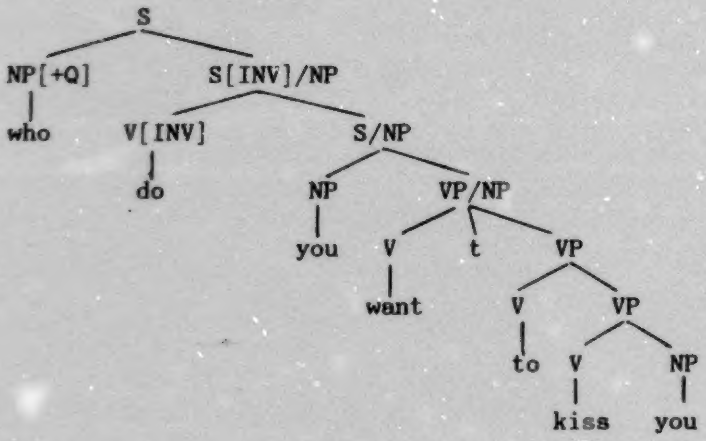
CC is indeed possible. Thus the sentence in (3.75) has the same interpretation as the version of the unreduced sentence in which the object in the embedded clause is what is topicalized, i.e. the I want to succeed Teddy reading.

Given the analysis of root wh questions discussed in the preceding section, a similar treatment is available for sentences like (2.24b) and (2.25b). The structures corresponding to these sentences are shown in (3.78):

3.78) a.



b.



Since the verb want in (3.78a) immediately precedes to and is also its aunt CC is possible and, thus (2.24b) is grammatical. CC is blocked for (2.25b), however, because of the intervening trace shown in (3.78b). Notice that unlike recent transformational accounts, there is no need to distinguish here between different types of traces (i.e. case marked versus non-case marked) since the rules responsible for wh sentences do not reapply for each S (i.e. are not successive cyclic) and thus do not overgenerate null elements.

All of the cases of unbounded dependency discussed thus far have involved structures in which an element which shares the category value of the SLASH feature is linearized to the left of the category that bears this feature. English, however, also allows sentences with rightward dependencies, an example of such a sentence being the one in (3.79b) in which an object NP is shifted to the end of its verb phrase:

3.79) a. I want all of the students who failed the exam to report to my office.
b. I want to report to my office all of the students who failed the exam.

Another possible treatment of sentences like (3.79) would be to allow the LP statements to fail to order objects and complements with respect to one another if the object is "heavy". Since heavy objects and complements would not be ordered, both (3.79a) and (3.79b) would be admissable linearizations of the verb phrase ID rule. There are, however, a number of arguments against this alternate analysis.

Aside from weakening the theory by allowing LP statements more power, a linearization analysis would remove the structural basis for the characteristic intonation pattern of rightward dependencies. More importantly, however, sentences like (3.80) could not be accounted for straightforwardly:

3.80) I want to report to my office and will speak to personally all of the students who failed the exam.

If sentences like (3.79b) are produced via linearized ID rules such as VP -> V VP NP, the verb want and the VP to report to my office would not form their own constituent. Thus, we would not expect them to operate as a unit for the purposes of conjunction, as they do in (3.80). If, however, what we have here is a rightward dependency, sentences like (3.80) can be produced; they simply involve the conjunction of two VP/NP's, want to report to my office and will speak to personally. One could of course claim that (3.79b) and (3.80) are produced by different means, (3.79b) via linearization and (3.80) with a rightward dependency rule, but such a move unnecessarily complicates the grammar since rightward dependency alone is sufficient to account for both.

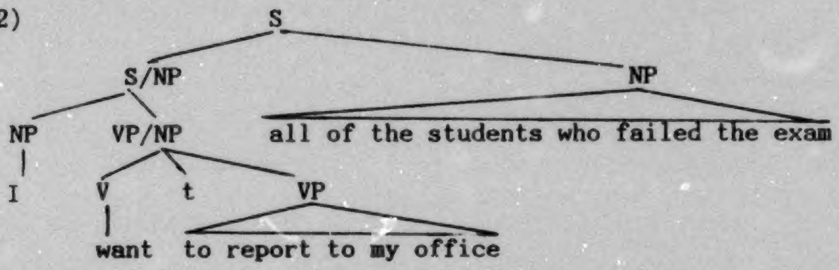
A further argument against a linearization analysis is provided by sentences in which the object appears to have been shifted outside of its VP:

3.81) I wanted to report to my office yesterday all the students who failed the exam.

On the interpretation where it was yesterday that I wanted X to happen, the adverb yesterday will be modifying the entire sentence. We will therefore have an ID rule like S -> S, Adv. Since the NP object in (3.81) is outside of the sentential adverb, it must also be outside of its own VP. Thus a linearization analysis cannot account for the clause order in (3.81) under this interpretation.

Given that sentences such as (3.79b) involve dependencies, the rules we have discussed thus far will provide structures like the following, assuming the correct linearization principle[34]:

3.82)


S
S/NP
NP
NP
VP/NP
all of the students who failed the exam
I
V
t
VP
want
to report to my office

Since the LP statements will place the trace of the object NP between the want and to of the matrix VP this analysis correctly predicts that sentences like (3.79b) will not allow CC:

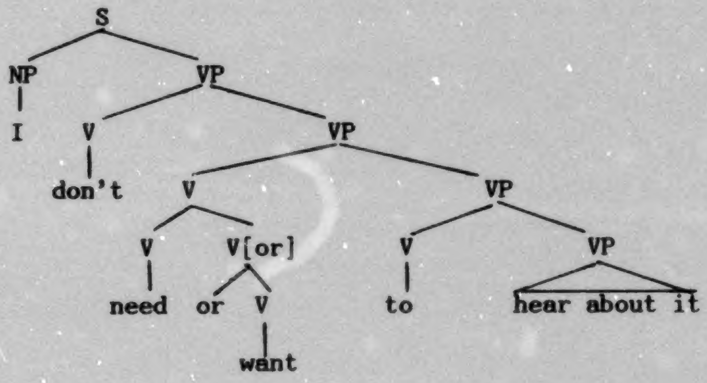
3.83) *I wanna report to my office all of the students who failed the exam.

This analysis of CC also predicts the failure of contraction in sentences with conjoined verbs such as (2.9a), repeated below:

2.9) a. I don't need or want to hear about it.

The structure assigned to such a sentence is shown in (3.84):

3.84)

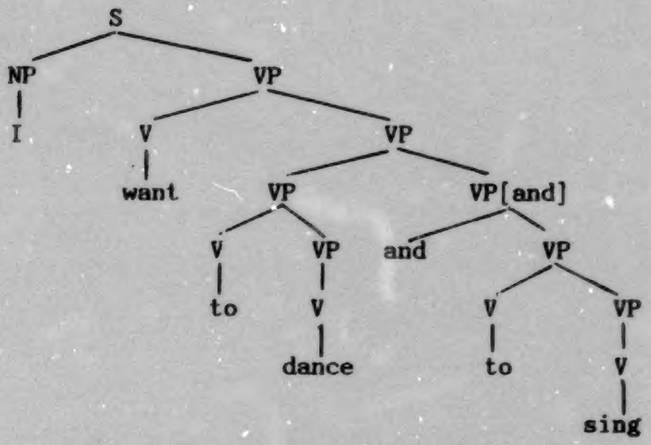

S
NP
VP
I
V
VP
don't
V
VP
V
V[or]
V
VP
need
or
V
to
hear about it
want

Since the V dominating want does not c-command the to, it is not its aunt and contraction is not possible there. Nor can contraction apply at the next higher level since that V does not introduce a trigger verb[35]. Since we are assuming that CC is not a syntactic rule we need not worry about possible ordering paradoxes between ID rules and/or metarules[36].

A similar treatment is possible for sentences like (2.8a) in which verb phrases are conjoined:

2.8) a. I want to dance and to sing.

3.85)

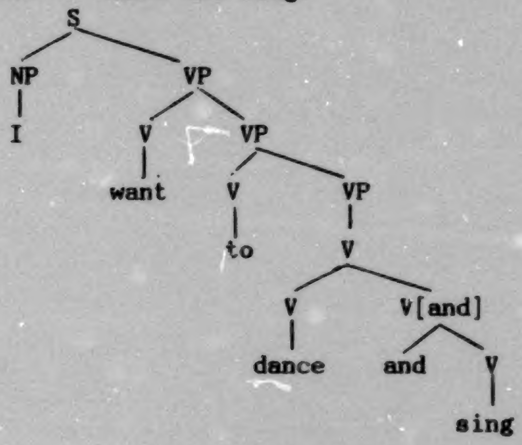

S
NP
VP
I
V
VP
want
VP
VP[and]
V
VP
and
VP
to
V
V
VP
dance
to
V
sing

Since the node dominating want is not the aunt of the node dominating the (first) to in this structure, CC is blocked. Thus sentences like (2.8b) are correctly predicted to be ungrammatical:

2.8) b. *I wanna dance and to sing.

Notice, however, that a sentence like (3.86a) will allow reduction since the structure it is assigned meets all the necessary requirements:

3.86) a. I want to dance and sing.
b.


S
NP
VP
I
V
VP
want
V
VP
to
V
V
V[and]
dance
and
V
sing

Thus the analysis presented above correctly handles the entire range of contraction and conjunction facts.

3.5.1 Liberal dialects.

Like AR, judgments about CC are subject to variation from dialect to dialect. Unfortunately the scope of this variation is not quite as well documented as with AR, thus making generalizations difficult. Some speakers seem to be more "liberal" than most with regard to CC in that the presence of trace between a to and its aunt does not block contraction. Such speakers find sentences such as (3.87) perfectly acceptable (though perhaps marked as very informal):

3.87) Who do you wanna kiss you?

Whether traces in such dialects behave differently with respect to AR as well has not, to my knowledge, ever been discussed; thus there is no way of telling whether the transparency in (3.87) is a fact about traces in general in these dialects or simply a fact about CC.

If the latter is the case then we need only change the condition on CC to something like the following: CC is possible in a configuration XYZ iff X is the aunt of Z and Y is phonologically null. Since the function of rules like CC and AR seems to be to make phonological units out of syntactically distinct items, it does not seem unreasonable that in some dialects the syntactic aspect would be the deciding factor in contraction while in others the phonological aspect would. Thus the two dialects differ only in minor details.

If, on the other hand, this transparency is a fact about traces as a whole for these speakers we have two options: we could either modify both AR and CC so as to allow traces in the relevant sites, or we could adopt a position for these dialects only similar to the one taken in GKPS and fail to have STM1 introduce traces. Which view of traces is correct and which modification of the theory is preferred is an empirical question in need of further research.

3.6 Ordering of AR and CC.

In section one I presented arguments against a morphological or phonological treatment of AR and CC as well as several preliminary arguments against analyzing either one as a syntactic rule. These latter arguments revolved around the fact that neither AR nor CC interacts crucially with any other syntactic rule and the fact that rules like AR and CC have very different functions than syntactic rules. The view of grammatical organization that I adopted there was one in which rules like AR and CC were contained in a separate component of the grammar reserved for cliticizations. While this type of highly constrained non-syntactic treatment of cliticization is consistent with most transformational analyses of AR and CC, it is in no way predicted by them. As we shall see, a non-syntactic treatment is not only consistent with the GPSG system it is actually required by it.

Consider first of all the type of statement a rule like AR would require if it were part of a GPSG syntax. As we saw in section (3.4), AR in the liberal dialect is possible in many different structures onto many different elements -- NP's, AdvP's, VP's etc. Thus, there will be no general way of specifying what the result of cliticization would look like. This plus the fact that rules in GPSG can in general only refer to two levels of structure at a time means that each of the eligible structures, if characterizable at all, would require a separate rule. This in turn suggests the possibility of dialects which contain some of these rules but not others, or which contain some rules with quite different constraints. Thus it would be logically possible for a dialect to have rules which allow AR in some cases only if the auxiliary is not followed by a phonologically non-null item and in other cases only if it is. The fact that the actual rules for AR in the dialects studied thus far all share the same constraint on following context would be treated as accidental. Since this is obviously not a desirable prediction we need some principled way of excluding this possiblility.

A second and even more damaging problem for a syntactic treatment of AR and CC in GPSG hinges on the fact that PS rules act as node admissibility conditions. Given this interpretation, it makes no sense to talk about derivations or ordered relationships among PS rules. Thus, a phrase is well formed if there is a PS rule for that phrase which allows the node it expands to dominate the categories it does in the proper order. This means that if AR or CC were syntactic rules the categories they introduced could be assigned structures by other rules in such a way as to place an undesirable element in the relevant position. In other words, there would be no way of preventing future instantiations of rules (whether basic or formed by metarule) from violating the conditions on cliticization.

For example, since clitics do not change the syntactic category of what they attach to, if a sentence like (3.87b) was formed on the basis of a sentence like (3.87a) the resulting form wanna would still be a verb.

3.87) a. I want to leave.
b. I wanna leave.

As such there would be no straightforward way of preventing this verb from being expanded by the conjunction rules responsible for sentences like (3.86). This would result in the generation of ungrammatical sentences such as (2.9b):

2.9) b. *I don't need or wanna hear about it.

Similarly, if a sentence like (3.88a) could be formed by the PS rules then a sentence like (3.88b) could also be formed by freely instantiating the rules responsible for (3.88a) with the feature [NUL]:

3.88) a. I am to leave and Pita's to leave also.
b. *I am to leave and Pitas's also.

We encounter the same type of difficulty with unbounded dependencies as well.

In order to maintain a syntactic analysis of AR and CC we would have to give up the view that PS rules are unordered as well as the idea that PS rules are node admissibility conditions, both of which are fundamental assumptions in GPSG. As a result we would be left with a less restrictive and much weaker theory. If we wish to preserve the theory as it is, we are forced to treat cliticizations as something distinct from syntax. The facts in a GPSG approach could not be otherwise without seriously altering its underlying claims. Thus we see that a theory of grammatical organization that has been argued for on independent grounds by many others falls out automatically if we employ a GPSG syntax.

Footnotes

*I would like to thank a number of people for their comments on various versions of this work. They are (in no particular order): David Dowty, Arnold Zwicky, Mike Geis, Brian Joseph, Rob Fox, Greg Stump and the members of Dowty's 1982 seminar on phrase stucture grammars. Any remaining errors are, of course, my mother's fault. I would also like to thank Karen Goldman and Isa Soto for kindly not finishing their theses before I finished mine. Erhard seems to think that he too deserves special mention.

1. Among others: Zwicky (1977), Klavans (1980), Zwicky (1982), Zwicky and Pullum (1982), Kaisse (1983a,b), and Pullum and Zwicky (forthcoming).

2. As developed in Gazdar (1981, 1982); Gazdar, Pullum and Sag (1982); Gazdar and Pullum (1981); Gazdar, Klein, Pullum and Sag (1982).

3. Klavans (1979) claims that another such distinction is that clitics always attach outside of any inflectional endings the host may have, arguing that apparent cases of endoclisis mentioned in Zwicky (1977) actually result from the clitic itself bearing suffixes.

4. Perlmutter (1970), George and Toman (1976), Klavans (1980).

5. Klavans argues that classical Greek provides examples of (1.7b) with so-called "stranded proclitics" such as ou in (i):

i. pôs gàr oú?
"for why not"

However, as Klavans herself points out (p.144) due to pre-pausal stress rules there is no way to tell if the "clitic" is truly attached to the following sentence. Since the element does have stress and does stand on its own, it is worth questioning in what sense it is a bound dependent in such sentences.

6. It should be noted that certain persons and tenses of these auxiliary verbs cliticize more freely than others. For the most part I will restrict myself to the forms is and has when discussing AR since they reduce most readily. In addition certain phonological considerations seem to discourage (though do not render impossible) AR. For a more complete discussion of the morphological and phonological factors involved in AR see Kaisse (1983a).

7. It should be pointed out that some dialects of English have more stringent restrictions on AR than others. These usually refer to the syntactic category of the phrase containing the host rather than the category of the word actually receiving the clitic. Therefore, even in conservative dialects, a variety of elements can serve as host.

8. An exhaustive list being: aspectual go, aspectual used, necessitative got, necessitative have, ought, suppose, and want in the sense of desire (rather than lack).

9. The fact that the [əz]/[ɨz] forms appear, at first glance, to have a wider distribution with reduced is/has than with the plural, third

person or possessive may simply be a result of there being phonological reduction rules with this output as well. Thus we have to distinguish between the two sources for these forms. In fact, in my own speech I prefer the vowelless alternate to the [əz]/[ɨz] forms at slow rates in sentences like:

i. John is going.
ii. Pete is going.

At faster rates [əz]/[ɨz] is acceptable, indicating that its ability to occur in positions not predicted by the general allomorphy rule under discussion is the result of a phononogical reduction rather than AR.

10. As argued for by Riemsdijk and Williams (1981).

11. While I find sentences like (2.16) perfectly grammatical there *are* sentences with Negative Constituent Presposing which do not sound as good:

i. ?Never's he to darken my door again.

I think this is because preposing with *never* is extremely stilted for me even *without* cliticization. Thus I would disagree with Kaisse's claim that there are no stylistic effects whatsoever involved in AR.

12. A possible response to this criticism would be that for purposes of AR the gap left by subject Wh-Movement "counts" as the first constituent in the sentence. Thus the reduced auxiliary in (2.22) and (2.3a,b) would still be a 2P clitic. However, such an analysis could not give a natural account of the presence of voicing assimilation in (2.22) or (2.3a) and would also reintroduce the notion of "gap" into Kaisse's analysis, something she had argued is not necessary.

13. This is not entirely clear from her article.

14. For a fuller discussion of the drawbacks to Selkirk's analysis see Pullum and Zwicky (forthcoming) and works cited therein.

15. This observation is due to Schmerling (unpublished manuscript). Note, however, that Postal and Pullum (1978) claim that a few speakers *do* accept sentences like (2.34b) (the one example they cite is Terry Langendoen), though they admit that they themselves find these sentences ill-formed. I am not aware of any other speakers who accept such sentences.

16. Sag and Klein also point out that rule numbers can be eliminated entirely in favor of indices on lexical categories. Furthermore, if you assume that metarules only apply to ID rules that introduce lexical categories and that numerical indices are contained in the feature matrices of lexical categories, then the claim that these indices are preserved under metarule application follows automatically. This is the approach adopted in some of the most recent GPSG articles.

17. See Gazdar and Pullum (1982) for a more complete description of the GPSG feature system.

18. In order to assure finiteness of the set of categories no feature is allowed to take itself as a coefficient (Gazdar and Pullum (1982)).

19. FOOT features are sometimes referred to as "binding features" as well.

20. Some versions of GPSG make use of a metagrammatical placeholder H_n, where 0 <= n <= 3, to represent the head daughter in an ID rule. The HFC insures the identity of features between H_n and its mother node. Gazdar and Pullum (1981) point out several advantages to this definition of head: 1) it allows the HFC to operate more generally in that it is responsible for all feature identity between mother and head daughter rather than just some of it, 2) the notion H simplifies the analysis of word compounds in English eliminating the need for parochial definitions of head in such cases and 3) H allows generalizations about English word order to be captured easily and without redundancies by the LP statements. For more details on the use of "H" see Gazdar and Pullum (1981, 1982).

21. This formulation of VP Deletion is slightly different from the one given in Gazdar, Pullum and Sag (1981) but more in keeping with the current approach to features. Since traces must retain their other category features there seems to be no need to have the trace dominated by VP[+NUL] as in earlier works.

22. This rule differs from the one given in GKPS in that it explicitly includes a trace element in its output.

23. In some current versions of GPSG this agreement is predicted by the Control Agreement Principle which says, simply, that verbs agree with their controllers. For more details see Klein and Sag (1982) and Gazdar and Pullum (1982).

24. These instantiations must also be consistent with any feature cooccurence restrictions (FCR's) which may apply, such as the FCR's which forbid a VP or A' from carrying a Wh feature.

25. Whether we want to give up this stipulation altogether is unclear. We may wish to have one class of metarules with this restriction and one without. If we do give it up entirely we will have to restrict the application of STM1 and STM2 appropriately, perhaps by reviving the Generalized Left Branch Condition.

26. Alternately one may wish to call the output of (3.45) something other than "S" to distinguish them from non-interrogatives; this is a minor detail.

27. This approach also entails a slightly different view of the organization of the grammar than the one taken by Sag and Klein outlined in section 3.1. Instead of ordering feature instantiation principles after metarule application, we must allow features to pass onto the ID doubles themselves in order to have S[+Q] expansions to serve as input to the rule in (3.45). Again, this is very similar to the treatment found in earlier versions of GPSG.

28. This version of the SAI metarule is based on the one given in Gazdar, Pullum and Sag (1981).

29. Given the organization of the grammar discussed in fn. 27 we would not even have to specify these features in the output of the SAI metarule. This is because 1) the HFC will insure that they are assigned to the V in the input rule and 2) features are preserved under metarule application unless otherwise specified.

30. The details of how such an FCR would work are not immediately relevant.

31. The discussion of so-called "Inversion" sentences presented here is greatly simplified and a number of details remain to be worked out. In particular, I avoid addressing the effects of definiteness and "heaviness" on linearization possibilities, cf:

i. a. Under this slab there is buried Joan of Arc.
   b. ?Under this slab there is buried Joan.
ii. a. Under this slab there is someone buried.
   b. *Under this slab there is Joan of Arc buried.

Such issues are beyond the scope of this work and in most cases do not critically affect the distribution of traces, which is our primary interest here. Also, it is not immediately clear whether we wish to maintain the distinction between existential there sentences, such as (3.66), and so called "presentational there" sentences, such as (3.67), suggested in Aissen (1975). These questions will be addressed in future work.

32. However one defines this class.

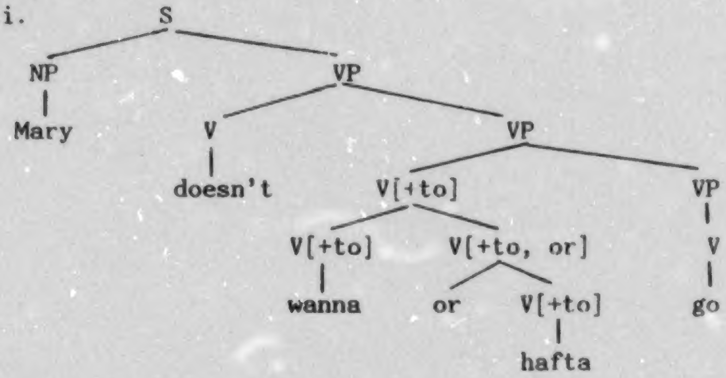
33. Again, I refrain from giving a formal rule for CC since it has been argued and will be argued again later that AR and CC are not and can not be syntactic rules in a GPSG system.

34. Exactly how these principles will be stated is somewhat problematic since there are certain restrictions on rightward dependencies that are not shared by leftward dependencies. Jacobson (1983 oral presentation, OSU) suggests that the non-unboundedness of rightward dependencies be captured by treating them as "double slashes" rather than single slashes, where double slash dependencies have the property of not being able to pass through certain types of nodes (bounding nodes). These issues, however, are outside the scope of this thesis; the matter that concerns us here is the placement of any trace elements, which would presumably be the same in both approaches.

35. Alternately, we could allow reduction to take place here with the result that the feature [+to] (or something to that effect) is assigned to the V. This feature will then trickle down onto each conjunct by general feature passing principles to produce sentences such as (i) with structures as in (ii):

i. Mary doesn't wanna or hafta go.

ii.



where a V[+to] is realized as its cliticized counterpart. Notice that the sentence in (i) will also have a possible source in:

iii. Mary doesn't want to or have to go.

Such an analysis can thus account for the feelings of some speakers that the sentence in (iii) requires "right node raising intonation" while the reduced sentence in (i) does not. Since (i) can also have the structure in (ii), it need not be analyzed as involving RNR. This is of course all quite speculative and depends upon adopting a particular view of what cliticization rules "look like". Whether this plan is feasible or not requires a great deal more study.

36. This point will be discussed further in the next section.

References

Aissen, J. (1975). "Presentational-There Insertion: A Cyclic Root Transformation." Chicago Linguistic Society 11, 1-14.

Anderson, P.K. (1979). "Word Order Typology and Prepositions in Old Indic", in Bela Brogyani (ed.) Studies in Diachronic, Synchronic, and Typological Linguistics. John Benjamin: Amsterdam.

Bresnan, J. (1971). "Contraction and the Transformational Cycle in English." mimeo. [IULC reproduction 1978]

Chomsky, N. (1976). "Conditions on Rules of Grammar." Linguistic Analysis 2, 303-351.

Chomsky, N. (1977). "On Wh-Movement", in P. Culicover, T. Wasow, and A. Akmajian (eds.) Formal Syntax. Academic Press: New York.

Chomsky, N. (1980). Rules and Representations. Columbia University Press: New York.

Chomsky, N. and H. Lasnik (1977). "Filters and Control." Linguistic Inquiry 8, 425-504.

Emonds, J.E. (1976). A Transformational Approach to English Syntax. Academic Press: New York.

Gazdar, G.J.M. (1981). "Unbounded Dependencies and Coordinate Structure." Linguistic Inquiry 12, 155-184.

Gazdar, G.J.M. (1982). "Phrase Structure Grammar", in P. Jacobson and G.K. Pullum (eds.) The Nature of Syntactic Representation. D. Reidel: Dordrecht.

Gazdar, G.J.M., E. H. Klein, G.K. Pullum and I.A. Sag (1982). "Coordinate Structure and Unbounded Dependencies". Stanford Working Papers in Grammatical Theory vol. 2. IULC: Bloomington

Gazdar, G.J.M. and G.K. Pullum (1981). "Subcategorization, Constituent Order, and the Notion "Head"", in T. Hoekstra, H. v.d.Hulst, and M. Moortgat (eds) Lexical Grammar II. Foris Publications: Dordrecht.

Gazdar, G.J.M. and G.K. Pullum (1982). Generalized Phrase Structure Grammar: A Theoretical Synopsis. IULC: Bloomington.

Gazdar, G.J.M., G.K. Pullum and I.A. Sag (1982). "Auxiliaries and Related Phenomena." Language 58, 591-638.

George, L. and J. Toman (1976). "Czech Clitics in Universal Grammar." Chicago Linguistic Society vol. 12, 235-249.

Jacobson, P. (1982). "Evidence for Gaps", in P. Jacobson and G.K. Pullum (eds) The Nature of Syntactic Representation. D. Reidel: Dordrecht.

Kaisse, E.M. (1983a). "The Syntax of Auxiliary Reduction in English." Language 59, 93-123.

Kaisse, E.M. (1983b). "The English Auxiliaries as Sentential Clitics", in Papers from the Parasession on the Interplay of Morphology, Phonology, and Syntax. Chicago Linguistic Society: Chicago.

King, H.V. (1970). "On Blocking the Rules for Contraction in English." Linguistic Inquiry, 134-136.

Klavans, J. (1979). "On Clitics as Words." Chicago Linguistic Society 15.

Klavans, J. (1980). Some Problems in a Theory of Clitics. Ph.D dissertation. University College London.

Klein, E.H. and I.A. Sag (1982). "Semantic Type and Control." Stanford Working Papers in Grammatical Theory vol. 2. IULC: Bloomington.

Lakoff, G. (1970). "Global Rules." Language 46, 627-639.

Lakoff, G. (1972). "The Arbitrary Basis of Transformational Grammar." Language 48, 76-87.

Nida, E.A. (1946). Morphology: The Descriptive Analysis of Words. University of Michigan Press: Ann Arbor.

Permutter, D. (1970). "Surface Structure Constraints in Syntax." Linguistic Inquiry 1, 187-256.

Postal, P.M. and G.K. Pullum (1978). "Traces and the Description of English Complementizer Contraction." Linguistic Inquiry 9, 1-29.

Pullum, G.K. and P.M. Postal (1979). "On an Inadequate Defense of Trace Theory." Linguistic Inquiry 10, 689-706.

Pullum, G.K. and P.M. Postal (1982). "The Contraction Debate." Linguistic Inquiry 13, 122-138.

Pullum, G.K. and A.M. Zwicky (forthcoming). The Syntax-Phonology Interface.

Riemsdijk, H. and E. Williams (1981). "NP-Structure." Linguistic Review 1:2, 171-217.

Sag, I.A. and E. Klein (1982). "The Syntax and Semantics of English Expletive Pronoun Constructions." Stanford Working Papers in Grammatical Theory vol. 2. IULC: Bloomington.

Selkirk, E.O. (1972). The Phrase Phonology of English and French. PhD dissertation. MIT.

Wood, W. (1979). "Auxiliary Reduction: A Unified Account." Chicago Linguistic Society 15, 366-377.

Zwicky, A.M. (1977). On Clitics. IULC: Bloomington.

Zwicky, A.M. (1982). "An Expanded View of Morphology in the Syntax Phonology Interface." Preprints for the XIIIth International Congress of Linguists, Tokyo.

Zwicky, A.M. and G.K. Pullum (1982). Cliticization versus Inflection: English n't. IULC: Bloomington.

OSU WPL 31 (1985) 122-129

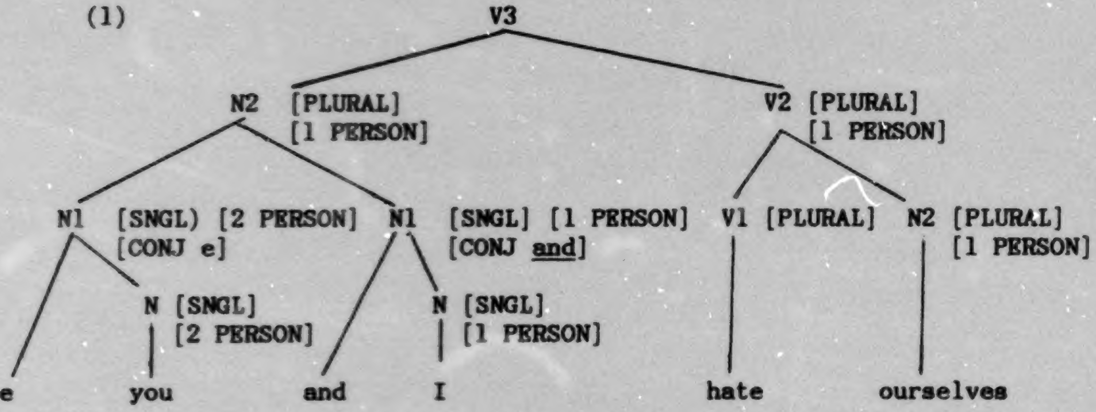
A New Approach to Feature Instantiation in GPSG*

Erhard W. Hinrichs

I. Introduction

It has been known for quite some time that the instantiation of agreement features on conjoined NPs presents a challenge for any adequate theory of agreement. Unlike other syntactic categories, for which agreement features in coordinate structures distribute from the mother node onto each conjunct, conjoined NPs can require non-identity between features of the mother and the features of one or more conjuncts. A notorious case in English concerns the conjunction and which typically requires plural agreement on the mother, even if all conjuncts have singular agreement. Moreover, the instantiation of agreement features for number is dependent on the following implicit hierarchy of feature values for number: [1 Person] > [2 Person] > [3 Person]. If two or more conjuncts differ in their values for person, the mother node will inherit the feature of that conjunct whose feature value is highest on the person hierarchy. These facts about person and number agreement concerning and are illustrated by the coordinate structure in (1).[1]

(1)



That the subject NP is in fact plural and first person can be derived from the first person plural form of the reflexive pronoun is object position.

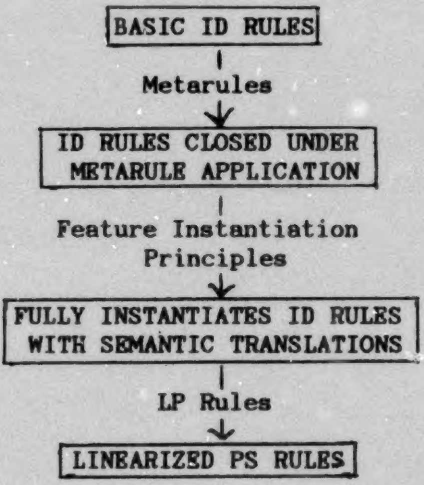
Crosslinguistically, non-identity of features in conjoined NPs is not restriced to person and number, but can also involve gender as in the French example in (2)[2].

(2) Un savoir et une adresse merveilleux
'a knowledge (MASC) and a skill (FEM) marvellous (MASC PLURAL)'
a marvellous knowledge and skill

In this paper I do not address the general issue of covariance of agreement features in conjoined NPs, but rather discuss an interesting subcase of this more general problem, namely agreement patterns of conjoined NPs which depend on the linear order of the conjuncts involved. In particular, I am concerned with agreement between verbs and conjuncts nearest to the verb in a following coordinate NP. Although this phenomenon has been pointed out with respect to English--compare the examples in (3) taken from Bach (1983)--I will mainly concentrate on examples from German and Russian because of the richer morphology and the greater variability of word order in these languages. Moreover, in Russian the conjunction 'i' 'and' can appear in front of every conjunct in a coordinate structure which will prove to be an important detail when we discuss the ramifications of agreement controlled by the conjunct nearest to the verb for the process of feature instantiation in Generalized Phrase Structure Grammar (GPSG). I will assume familiarity with the theory of GPSG as characterized in Gazdar/Pullum (1982). The basic organization of the theory as outlined in that paper can be summarized as in (4).

(3) a. There was a detective and three policemen in the room.
b. There were three policemen and a detective in the room. (Bach 1983,83)

(4)


BASIC ID RULES
Metarules
ID RULES CLOSED UNDER METARULE APPLICATION
Feature Instantiation Principles
FULLY INSTANTIATES ID RULES WITH SEMANTIC TRANSLATIONS
LP Rules
LINEARIZED PS RULES

The schema in (4) shows that in a GPSG as defined in Gazdar/Pullum (1982) feature instantiation properly precedes the linearization of syntactic constituents. My main claim in this paper is that such an ordering cannot be maintained if one wants to account for agreement controlled by the first conjunct of a coordinate structure. In order to account for such agreement patterns, one should rather conceive of feature instantiation principles as a set of wellformedness constraints on linearized and semantically translated PS rules. But before I can make this alternative proposal more precise, let me present the relevant data in German and Russian.

II. The Data

According to Drach's Law the finite verb in German declarative clauses occupies the second position in the sentence. Usually the first constituent is the subject, but it can also be fronted prepositional phrase as in (5), an adverbial as in (6), or the dummy es as in (7).

(5) a. Zu der Sitzung kamen (3PL) die neue Professorin
'To the meeting came the new professor
und alle Assistenten.
and all assistants.'

b. Zu der Sitzung kam (3SG) die neue Professorin (3SG)
und alle Assistenten.

c. *Zu der Sitzung kam (3SG) alle Assistenten (3PL) und
die neue Professorin (3SG).

(6) a. Nächste Woche könnt (2PL) Fritz und Du uns besuchen.
'Next week may Fritz and you us visit.'

b. Nächste Woche kann (3SG) Fritz (3SG) und Du uns
besuchen.

c. *Nächste Woche kann (3SG) Du und Fritz (3SG) uns
besuchen.

(7) a. Es protestierten (3PL) die Fraktion der SPD und die Grünen.
'It protested the faction of the SPD and the
Greens.'

b. Es protestierte (3SG) die Fraktion der SPD (3SG) und die
Grünen.

c. *Es protestierte (3SG) die Grünen (3PL) und die Fraktion
der SPD (3SG).

If the subject NP is a conjoined NP as in (5)-(7), finite verb can either agree with the conjunct nearest to the finite verb, or with the conjoined NP as a whole. In the latter case the verb will always be marked as plural, as in (5)a-(7)a. But if the agreement is controlled by the first conjunct and if that conjunct is marked as singular, the finite verb is singular, as in (5)b-(7)b. Agreement with any one conjunct is restricted to the nearest conjunct only, as the ungrammaticality of (5)c-(7)c shows.

The same phenomenon can be found in Russian. If the verb precedes a coordinate structure with the conjunction 'i'meaning 'and', the verb can either agree with the coordinate structure as a whole, which is exemplified in sentences (8)a and (9)a, or the verb can agree with the nearest conjunct, which is the case in sentences (8)b and (9)b.

(8) a. Prepodavalis' (PL) čerčenie i matematika.
'Was taught graphics and mathematics.'

b. Prepodavalos' (Neut SG) čerčenie (Neut SG) i
matematika.

(9) a. Na sobranie prišli (PL) professor i pjat' studentov.
'To the meeting came the professor and five students.'

b. Na sobranie prišel (Masc SG) professor (Masc SG) i
pjat' studentov.

For the purposes of this paper I will only analyse coordinate stuctures which appear as subject NPs following the verb. Corbett (1982) claims that in Russian agreement controlled by the conjunct nearest to the verb is also possible if the subject coordinate structure precedes the verb. Under certain circumstances this is also possible in German, but because of interference by other factors such as semantic salience of the conjuncts involved, these data have to await further study.

III. Theoretical Significance of the Data

If one wants to account for the agreement facts in German and Russian in a GPSG as outlined in (4), the following problem arises. Since agreement features are instantiated on the basis of unordered constituents, there is *a priori* no guarantee that the daughter constituent matching the agreement features on the mother will be the conjunct nearest to the verb. If one wanted to maintain the overall organization of GPSG outlined in (4) and thus apply feature instantiation to unordered constituents, one would have to distinguish the conjunct controlling agreement from all others by means of some special syntactic feature label. Of the categories used in GPSG, the one that comes to mind, of course, is the head feature, especially since we are dealing with a case of identity between agreement features which are, after all, head features. And identity between head features is commonly handled by the Head Feature Convention. Thus, one might propose a PS rule as in (10) to generate coordination structures like *du und deine Freunde* in the German sentence in (11).

(10) N [BAR 2] --> H [CONJ e], N [BAR 2] [CONJ und]$^+$

(11) Natürlich kannst (2SG) Du (2SG) und deine Freunde bleiben.
'Of course can you and your friends stay.'
You and your friends can stay, of course.

The linearization rule in (12) would further guarantee that the conjunct controlling agreement, i.e. the head conjunct, will precede the non-head conjuncts.

(12) H [BAR 2] [+CONJ] < N [BAR 2] [+ CONJ]

The first problem for this type of approach, which I will refer to as the "Head Daughter analysis", concerns the number of ID rules and linearization rules that have to be stated separately, if agreement features are instantiated on unordered constituents for a language such as German. In addition to a general coordination schema as in (13), which is modelled after the schema proposed by Gazdar/Klein/Pullum/Sag (1982) for English, the rule in (10) has to be posited along with the linearization statement in (12) to account for coordinate structures controlled by the first conjunct. Moreover, in order to account for coordinate structures with the same agreement pattern, but with *und* appearing before the last conjunct only, a third PS rule as in (14) would have to be stated.

(13) a. CAT' --> CAT'[CONJ α_0], CAT'[CONJ α_1]$^+$

b. α is in {<und,e>, <e,und> <weder,noch>, <oder,e>, <e,oder>,<sowohl,als auch>}

(14) N [BAR 2] --> H [BAR 2] [CONJ e] , N [BAR 2] [CONJ e]$^+$, N [BAR 2] [CONJ und]

Notice that the number of rules necessary to generate all the relevant coordinate structures whose first conjuct agrees with a preceding VP would have to be even greater for a language like Russian. In additionn to the ID rules in (10) and (14), in which und would be replaced by i, we would need a rule as in (15), since in Russian all conjucts may be preceded by i.

(15) N [BAR 2] --> H [BAR 2] [CONJ i] , N [BAR 2] [CONJ i]$^+$

The second, and main objection to the Head Daughter analysis follows from the first one. Because such an analysis forces us in the case of German to use three separate rules in addition to a generalized rule schema for coordinate structures, the resulting grammar misses a number of significant generalizations. Unlike the analysis that I will present below, the head daughter analysis fails to treat coordination as a unified phenomenon by means of one generalized rule schema as in (13), but has to state three separate, and partially redundant PS rules. Moreover, even NP conjunction cannot be treated as a single phenomenon because two distinct ID rules are needed for the distribution of the lexical item und.

Furthermore, by disassociating the LP rule in (12) from the ID rules (10) and (14), it treats the linearization of constituents and the agreement pattern of conjoined NPs as logically independent, when, in fact they are crucially related. Because the rules are independent, the analysis suggests that there might be languages that do have ID rules like (10) and (14), but instead of (12) an LP rule requiring the head to alwyas appear as, say, the third conjunct. But to my knowledge, no such language exists and for perceptual reasons is unlikely to exist.

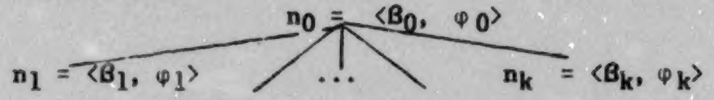
If, on the other hand, we allow feature instantiation principles to operate on linearized and semantically translated constituents, we can generate all relevant coordinate structures in German by just one generalized schema as in (11). To account for agreement controlled by the first conjunct, we only have to state one additional feature instantiation constraint, regardless of the distribution of the lexical item und in coordinate structures.

Let us briefly outline this alternative approach to feature instantiation. At the heart of my proposal is a one-to-one and onto mapping from constituents of PS rules specifying immediate dominance relations only to nodes of locally ordered trees whose nodes consist of ordered pairs of a syntactic category β and the semantic translation φ of the syntactic expression dominated by β.

(16) For each ID Rule

$\alpha_0 \longrightarrow \alpha_1, \ldots, \alpha_k$

and ordered tree


$n_0 = \langle \beta_0, \varphi_0 \rangle$
$n_1 = \langle \beta_1, \varphi_1 \rangle$
...
$n_k = \langle \beta_k, \varphi_k \rangle$

there is a one-to-one and onto mapping

$f : \{\alpha\}_i \longrightarrow \{n\}_i$ for $o < i < k$

such that $f(\alpha_0) = n_o$ and $\forall \alpha_i [f(\alpha_i) = \langle \beta_j\ \varphi_j \rangle \Longrightarrow \beta_j = \mathrm{EXT}(\alpha_i)]$

Function f in (16) maps the mother constituent of the ID rule onto the root of the tree. f has to be one-to-one so that every constituent of the ID rule is mapped onto a distinct node of thre tree, and vise versa. f also has to be onto so that every node of the tree is linked with some constituent of the corresponding ID rule. Moreover, each syntactic category β has to be an extension of the corresponding constituent of the ID rule.

Feature instantiation principles, under this approach, can be viewed as wellformedness constraints on possible mappings f from constituents of ID rules to nodes of trees. The Head Feature Convention, for example, can be defined as in (17).

(17) HFC: $\forall \alpha_i\ [\alpha_i = \mathrm{H}\ [\mathrm{BAR}\ n] \wedge f(\alpha_i) = n_j \Rightarrow \mathrm{HEAD}(\beta_j) = \mathrm{HEAD}\ (\beta_0)\ ]$ for $i,j \geq 1$.

Likewise, LP rules can be conceived of as constraints on the set of possible mappings between ID rules and ordered trees. The LP rule for English that requires lexical heads of major syntactic categories to precede sister constituents can be stated as in (18).

(18) $\forall \alpha_i, \alpha_j\ [\ \alpha_i, = \mathrm{H}\ [\mathrm{BAR}\ 0] \wedge \alpha_j \neq \alpha_i \Rightarrow f(\alpha_i) < f(\alpha_j)\ ]$ for $i,j \geq 1$.

Presupposing a mapping as in (16), let us turn to an analysis of German coordinate structures controlled by the first conjunct following the verb. I follow Uszkoreit (1982) and Nerbonne (1983) whose GPSG analyses of German account for Drach's Law by treating the first constituent in German declarative clauses as the result of topicalization. Subject NPs following the verb in second position can hence be identified by the features [-TOPICALIZED] and by [+ NOMINATIVE] as the value for the case feature. In order to generate non-topicalized coordinate subject NPs whose agreement features match that of the leftmost conjunct we have to impose the constraint (19) on possible mappings f from ID rules to ordered PS markers defined in (16).

(19) $\forall \alpha_0, \alpha_i\ [\alpha_0 = \mathrm{EXT}(\mathrm{N}\ [\mathrm{BAR}\ 2][-\mathrm{TOP}]\ [+\mathrm{NOM}]) \wedge f(\alpha_1) = n_1 \wedge \exists \alpha_j = [\alpha_j = \mathrm{EXT}([\mathrm{CONJ}\ und])] \Rightarrow \mathrm{AGR}\ (\beta_1) = \mathrm{AGR}(\beta_0)\ ]$, where $i,j \geq 1$.

The condition in (19) specifies that for any ID rule in which the dominating category is an extension of the feature N" [-TOP] [+NOM], and which has at least one daughter with the feature CONJ und, a daughter can only be mapped onto the leftmost node in such ordered trees where that node has the same agreement features as the root of the tree.

IV. Conclusion

By treating LP rules and feature instantiation principles as constraints on possible mappings from ID rules to linearized and semantically translated PS rules we arrive at an organization of GPSG in which LP rules on the one hand, and feature instantiation principles and semantic translation principles on the other hand apply in tandem, rather than inseparate components of the grammar. Such an organization of GPSG has been independently proposed by Klein and Sag (1982) to capture significant generalizations about grammatical relations and word order in English. While their work concerns the relationship between LP statements and semantic translation principles, the argument presented in this paper rests on the interaction between feature instantiation principles and LP statements.

FOOTNOTES

*I would like to thank Annie Bissantz, Remo Pereschi, and especially Ewan Klein and Arnold Zwicky for their helpful comments and suggestions on earlier drafts of this paper. I am grateful to Anelya Rugaleva for the Russian examples.

[1] The tree structure in (1) is modelled after the account of coordination given in Gazdar/Klein/Pullum/Sag (1982).

[2] The example in (2) is due to Corbett (1983).

BIBLIOGRAPHY

Bach, Emmon (1983). "On the Relationship between Word-Grammar and Phrase-Grammar". Natural Language and Linguistic Theory, 1, 65-89.

Corbett, Greville (1983). Hierarchies, Targets, and Controllers: Agreement Patterns in Slavic, London: Croom Helm.

Gazdar, Gerald and Geoffrey K. Pullum (1982). Generalized Phrase Structure Grammar: a Theoretical Synopsis. Distributed by Indianna University Linguistics Club.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag (1982). "Coordinate Structure and Unbounded Dependencies". M. Barlow, D. Flickinger and I. A. Sag eds. Developments in Generalized Phrase Structure Grammar: Stanford Working Papers in Grammatical Theory, Volume 2, 38-68.

Klein, Ewan, and Ivan A. Sag (1982). "Semantic Type and Control". M. Barlow, D. Flickinger and I. A. Sag eds. Developments in Generalized Phrase Structure Grammar: Stanford Working Papers in Grammatical Theory, Volume 2, 1-25.

Nerbonne, John (1982). "Phantoms and German Fronting: Poltergeist Constituents". Paper presented at the Annual Meeting of the LSA held at San Diego, California. December 1982.

Uszkoreit, Hans (1982). "German Word Order in GPSG". In: D. Flickinger, M. Macken, and N. Wiegand eds. Proceedings of the First West Coast Conference on Formal Linguistics. Stanford University, Stanford, California.

OSU WPL 31 (1985) 130-159

The Syntax of Conditional Sentences

Michael L. Geis
Department of Linguistics
The Ohio State University

Introductory Remarks

Although English has a number of different types of adverbial clauses, as is illustrated by such sentences as

(1) a. I will leave when you do.

b. I will go where you go.

c. I will leave before you do.

d. I will leave if you do.

e. I left because you did.

f. I left although I wasn't supposed to.

conditional clauses (see (1d)) have attracted the most attention from philosophers and linguists. Indeed, both linguists and philosophers have devoted whole conferences to their consideration. What is interesting about this from a linguistic perspective is that conditional clauses are not all that much more interesting linguistically than are any of the other types.[1] The reason for this special interest is surely that conditional sentences play a key role in reasoning, at least the sort of reasoning that interests philosophers.

Interestingly, not only have those who have studied conditonal sentences usually not discussed them in the context of other adverbial clauses, they have focused their attention almost exclusively on sentences like (1d) which employ the adverb *if*. Surely, however, conditional sentences in which *if* is modified by *only* and *even*, as in (2) and (3), are also of interest.

(2) I will leave only if you do.

(3) I will leave even if you do.

Moreover, there are totally ignored conditional constructions, very close in meaning to those just cited, which, like *if*-clauses, are adverbial in character:

(4) I will leave in the event that you do.

(5) I will leave only in the event that you do.

(6) I will leave even in the event that you do.

It is interesting to speculate on why sentences like (4)-(6) have been ignored. One possibility is the stylistic preference of logicians for the word _if_.[2] However, the main reason is surely that monomorphemic realizations of a category are typically perceived by native speakers (including philosophers) as its most basic representatives.[3]

In this paper, I present a syntactic analysis of a wide range of conditional sentences, which develops ideas first published in Geis (1973). A companion semantics and pragmatics for this analysis has been provided by Lycan (1984).[4] Since this study has been published, I will presume familiarity with it. What I propose to do here is provide the syntactic argumentation for this analysis, as well as the details of its formalization.

What one takes to be a conditional sentence will, I think, depend on whether or not one takes a syntactic or semantic perspective. Viewed semantically, a sentence like (7a) might be said to be conditional on the grounds that it has the "same" meaning as (7b), which clearly is conditional.

(7) a. Kiss my dog and you'll get fleas.

b. If you kiss my dog, you'll get fleas.

On the other hand, (7a) is not conditional in form, so it would be perfectly reasonable not to include this sentence in a study of conditional _sentences_, as opposed to conditonal _propositions_. Similarly, one might, following Stump (1981), take a sentence like (8) to be pertinent to a study of conditionals.

(8) For you to do that would be nice.

Certainly, (8) is conditional in meaning. However, again, there is little linguistic motivation for including such a sentence in the analysis of conditional _sentences_, though obviously it is relevant to the analysis of conditional _propositions_, i.e. of nonlinguistic mental sentences. I say this because (8), unlike (1d) and (2)-(6), is not conditional in form, and it is the possession of linguistic form that distinguishes real sentences from mental sentences (i. e. propositions). The fact that a sentence might be conditional in meaning does not qualify it for membership in the class of sentences containing conditional clauses. The reason is that we may look to extralinguistic semantic theories for an account of how it is that English has several different ways of expressing conditionality. As another example of this point, we might note that the fact that a pair of sentences like (9a) and (9b) might both express causality does not qualify (9a) as relevant to a study of causal clauses.

(9) a. John's leaving precipitated Bill's departure.

b. John left because Bill departed.

Too few linguists and philosphers seem to recognize that most putatively substantive claims about the relationship between syntax and semantics have been largely definitional in character. The once widely heralded claim (Harman 1972) that Deep Structure is Logical Form, is, perhaps the best example of this, but there are others.[5] To insist that sentences like (7a) and (8) must be brought into the picture in the attempt to describe the syntax of conditional sentences is, quite simply, to beg one of the most important questions of syntax: What is the contribution of syntax to the use and understanding of the sentences of our languages? The fact that native speakers of English "know" that (7a) and (7b) have the same meaning or "know" that (8) is conditional in meaning, taken alone, is not necessarily relevant to a linguistic analysis of conditional sentences, for speakers of English know more than just English. They, presumably, can do some elementary reasoning with sentences.[6]

If it is to be at all general, a linguistic description of English conditional sentences will want to account, at the very least, for such sentences as (1d) and (2)-(6), for these sentences are all conditional in form. It should perhaps also account for how these sentences are related to other adverbial clause constructions, such as those of (1), for these are also similar in structure. To my knowledge, the only comprehensive generative account of adverbial clauses in English is in Geis (1970a). However, Heinamaki (1974) and Larson (mss) have worked on temporal clauses, and Bresnan and Grimshaw (1978) on a similar construction.[7] I shall bring each of these studies into the picture as they become relevant. In Geis (1970a), conditional clauses were sharply distinguished from adverbial clauses introduced by _when_, _while_, and _where_ and by _before_, _after_, _until_, and _since_. Clauses introduced by this latter array of words were said to be a species of relative clauses. I argued (Geis 1970a), in particular, that adverbial subordinate clauses introduced by the above connectives are derived transformationally from underlying syntactic structures in which the clauses introduced by these words are explicitly relative in character. According to this view, a sentence like (10) was said to be derived from the structure underlying (11) by a rule of Antecedent Deletion.

(10) I will leave at the time when you leave.

(11) I will leave when you leave.

Conditional clauses, for reasons to be identified later, were said not to be relative clauses, but, instead, to be a species of nominal complements. This line was continued in Geis (1973). What I shall argue now is that _if_-clauses like _when_-clauses are themselves a species of relative clauses, as is implied by Lycan's semantics of conditionals. However, my present analysis, unlike the earlier transformational account of adverbial relative clauses, can be stated wholly in terms of English surface structures. On the analysis to be presented here, which is monostratal, I shall argue that a sentence like (12) has essentially the same sort of structure as (11).[8]

(12) I will leave if you leave.

The Adverbial Analysis of Conditionals

The Sentence Operator Version

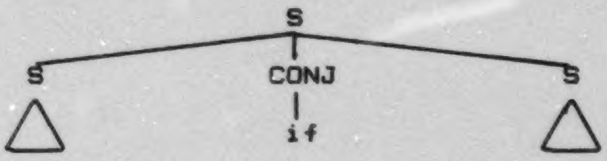
In the Propositional Calculus, a sentence having the form of (13) is usually assigned a representation like (14).

(13) If S1, then S2

(14) S1 $\supset$ S2

where (14) is understood to be false if S1 is true and S2 is false and is otherwise true.[9] According to this analysis, the two clauses that make up a conditional sentence are coordinated semantically as they would be syntactically if *if* were a "true conjunction," to borrow a phrase from Jespersen (1961: V.4.344).[10] On this view, sentence (1d) would be given an analysis something like (15).

(15)



Interestingly enough, despite the long tradition associated with the standard truth functional analysis of *if*, there is no solid syntactic evidence whatever that supports the division of conditional sentences into two coordinate sentences, as in this analysis.[11] It might be argued that *if*...,*then*... is structurally parallel to *either*...*or*... and *both*...*and*.... The parallel is, however, an illusion. As pairs like (16) and (17) show, *then* need not occur for a conditional sentence to be grammatical, but of course, *and* and *or* are obligatory in compound sentences.

(16) If you leave, then I'll leave.

(17) If you leave, I'll leave.

One cannot say

(18) *Both John left, Bill left.

(19) *Either John left, Bill left.

It is *if*, not *then* that is the more fundamental constituent of conditional sentences.

Though logicians normally cite conditional sentences in the form of (16) or (17), a syntactician is likely to take a sentence like (20), to be the more "natural," for in (20), the *if*-clause is in "normal" adverbial position.

(20) I will leave if you leave.

Note that the two clauses of (20) are the reverse of what they are in (13), (16) and (17). In the latter sentences the condition precedes the consequence, but in (20), the reverse is true. As noted, I take post-verbal conditional clauses to be in "normal" word order for conditional sentences. Talmy (1976) makes the universal claim in connection with causal constructions that causes are subordinated to effects in the languages of the world, as in (1e) and (1f), repeated here as (21) and (22), respectively.

(21) I left because you did.

(22) I left although you did.

This is to say that languages (other than formal languages, of course) do not have adverbial clause constructions in which the consequent is subordinated to the antecedent, from which it would appear to follow that (20) is more basic than (13), (16), and (17). Moreover, not all languages even have the capacity to place conditional clauses in sentence-initial position, as in (16) and (17).[12]

The fact that a sentence like (20) might be more natural than (13), (16), and (17) does not, by itself, upend the binary sentence operator analysis. One need only define a reverse horseshoe, with appropriately revised truth-conditions, and reverse the two clauses. On this revision, the *if* of (15) would correspond exactly to the reversed horseshoe. However, the fact that *if*-clauses can occur both sentence-initially and post-verbally is itself rather good evidence that *if*-clauses are adverbial. Simple and complex time adverbials both have this freedom of occurrence, for instance:

(23) a. I will leave at noon.

b. At noon, I will leave.

(24) a. I'll leave when you leave.

b. When you leave, I'll leave.

Data like (23) and (24) strike at the heart of the coordinating conjunction analysis of conditionals, for patterns like these do not obtain in the case of coordinate structures:

(25) a. Joe will leave and Mary will stay.

b. *And Mary will stay, Joe will leave.

(26) a. Joe will leave but Mary will stay.

b. *But Mary will stay, Joe will leave.

We have quite clear evidence that _if_-clauses are adverbial and are subordinate to main clauses, not coordinated with them.

There is solid syntactico-semantic evidence that post-verbal occurrences of _if_-clauses (see (20)) are more basic than preposed _if_-clauses (see (16) and (17)). Observe that sentences (27)-(29) have essentially the same interpretation.[13]

(27) I think that I will leave if you leave.

(28) I think that if you leave, (then) I'll leave.

(29) If you leave, (then) I think I'll leave.

It is clear that (29) is not to be interpreted as stating that there is a conditonal relationship between the hearer's leaving and the speaker's thinking about leaving, contra its surface form. Instead, the speaker is saying that he thinks that the hearer's leaving will lead to his leaving. In the standard transformational idiom, we would account for this by saying that the _if_-clause of (29) is put there by an extraction rule, namely Adverb Preposing (recall (23) and (24)). Even in monostratal theories of syntax a sentence like (29) must be treated as the "marked" form.

Observe that when _if_-clauses are preposed, as in (28) and (29), a second conditional adverbial cannot occur in postverbal position:

(30) *I think that if you leave, I'll leave in that event.

(31) *If you leave, then I think I'll leave in that event.

Thus, we must have some way to exclude a conditional adverbial from post-verbal adverbial position, if an _if_-clause occurs clause- or sentence-initially. Within GPSG, this is achieved, of course, via the slash-category notation.[14]

(32) S ----> ADV[2] S/ADV[2] (#9)

Rule (32) stipulates that if an S begins with an adverb phrase then the S it is sister to must have an adverb phrase gap. Given rule (32), we can account for both (30) and (31), as well as (28) and (29).

The adverb preposing data reveals the hopelessness of the standard analysis of the structure of conditional sentences in logic texts on syntactic grounds, and, as a result of this failure, it is hardly surprising that the analysis fails semantically as well. Adverbial constructions normally involve quantification over something--times, places, events, etc, as is implicit in the relative clause treatment I shall be giving later.[15] Such a fact is quite telling against the truth functional account of *if*.

The failure of the standard logical treatment of conditonals is further revealed by the fact that *if*-clauses can be modified by *only* and *even*. Consider:

(33) a. I will leave only if you do.

b. I will leave even if you do.

(34) a. John works only when his back feels good.

b. John works even when his back hurts.

(35) a. *John works hard only and Bill works hard.

b. *John works hard even and Bill works hard.

As (34) indicates, *only* and *even* are quite comfortable modifying adverbial *when*-clauses. Note, though, that they do not modify nominal *when*-clauses:

(36) a. He asked me yesterday when I would leave.

b. *He asked me yesterday only when I would leave.

We have here the clearest possible evidence that *if*-clauses of the sort we are interested in are adverbial, for[16] note that nominal *if*-clauses aren't modifiable by *only* either:

(37) a. He asked me yesterday if I would leave.

b. *He asked me yesterday only if I would leave.

Certainly the view that *if* might be a conjunction is falsified by (33)-(35).

Very clear evidence that conditional clauses, when they occur postverbally, are constituents of verb phrases and are therefore adverbials is provided by data involving VP Deletion and *Do So*.[17] Observe that the place-holders, *<>* and *do so* of the following sentences, are interpreted as referring back to the underlined verb phrases of these sentences:

(38) a. I will *leave at noon* and Joe will <> too.

b. I will *leave at noon* and Joe will do so too.

(39) a. I will *leave under certain circumstances* and Joe will <> too.

b. I will *leave under certain circumstances* and Joe will do so too.

(40) a. I will *leave when you do* and Joe will <> too.

b. I will *leave when you do* and Joe will do so too.

(41) a. I will *leave if you do* and Joe will <> too.

b. I will *leave if you do* and Joe will do so too.

The most conservative interpretation of these data is that the place-holders in these sentences refer back to *constituents*, and thus that temporal and conditonal constituents of the left conjuncts of these sentences are constituents of verb phrases. Given this and the fact that these constituents are not noun phrases, all that is left for them to be is adverbials (= prepositional phrases, adverb clauses, or adverbs.)

The data we have considered so far involving preposed conditional clauses, modification by *only* and *even* and verb phrase ellipsis phenomena provides very strong evidence against the view that the conditional and main clauses of conditional sentences are coordinate and for the view that conditonal phrases and clauses, like temporal phrases and clauses, are adverbial constructions. These arguments carry over to the other types of conditional clauses identified earlier. Thus, *if*-clauses modified by *only* and *even* prepose as do complex prepositional phrases like *in the event that S*, whether or not they are modified by *only* and *even*. VP-Deletion and *Do so* tests show that all of these constructions are or can be constituents of predicates.

There also exists quite direct evidence that conditonal clauses are adverbial in character, evidence that closely links *if*-clauses to *in the event that S* constructions on the one hand and to adverbial relative clauses on the other. Consider the following pronominalization data:

(42) a. I will leave *if you leave*, and Joe will leave *then* too.

b. I will leave if you leave, and Joe will leave in that event too.

(43) a. I will leave when you leave, and Joe will leave then too.

b. I will leave when you leave, and Joe will leave at that time too.

In these sentences, we have clauses whose anaphoric reflexes are adverbials, either simple conditional and temporal adverbs or more complex prepositional phrases. In this respect, if-clauses and when-clauses act like explicit prepositonal phrases. Compare (44) and (45) with (42) and (43), respectively.

(44) a. I will leave in the event that you leave and Joe will leave then too.

b. I will leave in the event that you leave and Joe will leave in that event too.

(45) a. I will leave at the time that you leave and Joe will leave then too.

b. I will leave at the time that you leave and Joe will leave at that time too.

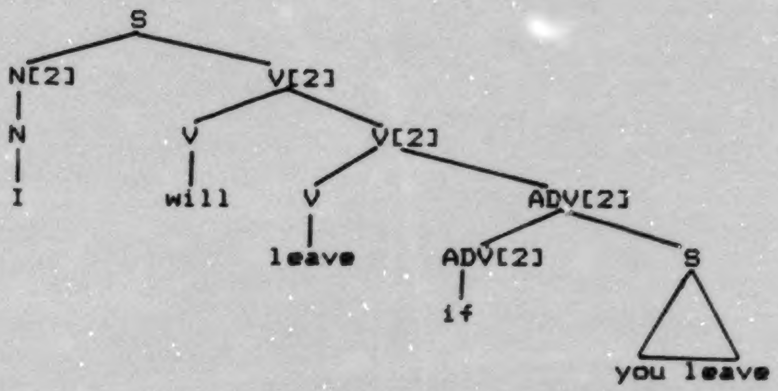
We would not want to conclude from these data that if-clauses and when-clauses modify explicit prepositonal phrases in underlying structure, as I once did in the fabulous days of Generative Semantics, but we are entitled to conclude from these data that if-clauses and when-clauses are adverbials.

There is an adverbial analysis of if-clauses that has a good deal of initial plausibility, namely one in which if is treated as a kind of sentence operator that turns ordinary clauses into adverbial clauses. On this view, if would be treated as in

(46) ADV[2] -----> ADV[+cond, +oper] S

According to this analysis, the structure of nonpreposed conditionals would look something like this:

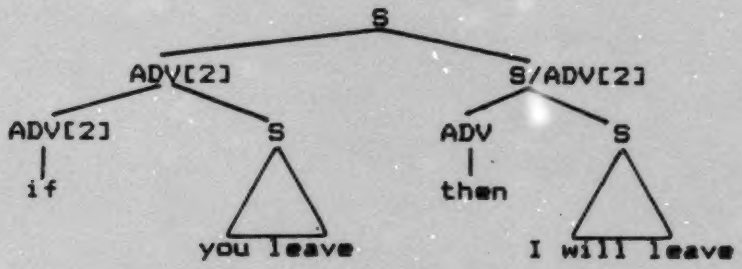
(47)



This analysis is appears to be consistent with at least some of the facts we have so far considered, and it provides a not at all implausible surface structure. However, there are troubling facts, not the least of which is the fact that *then* pops up in cases of preposed adverbials. Given the assumptions of (47), we would want, I think, to say that a sentence like (48) has a structure like (49).

(48) If you leave, then I will leave?

(49)



Though structure (49) is eminently reasonable, I believe that it is incorrect. There are two features of (49) that are somewhat problematic. There is nothing (explicit or implicit) in such an analysis that explains why *then* might pop up in conditional sentences. So it is potentially deficient on explanatory grounds. Moreover, there is evidence that *if* is a constituent of the clauses it introduces, contra (49).

The Relative Clause Analysis

The hypothesis that conditional clauses are adverbial in character is well-motivated, and the structures we have assigned to normal and preposed *if*-clauses are credible representations of the surface organization of such sentences. What is missing is a defense of the particulars of these analyses and some sort of explanation of the position of such structures in the grammar of the language as a whole.

Throughout our discussion of _if_-clauses, we have found that _if_-clauses and _when_-clauses parallel each other exactly: both prepose; both accept the modifiers _only_ and _even_; both are consituents of verb phrases when they occur postverbally; and both pronominalize the same way, to the point of sharing the homophonous adverbial proadverb _then_. This latter point is worth pursuing further. There are hosts of languages that employ temporal expressions exclusively to express conditionality. Even English, which has a plenitude of conditional expressions, allows its speakers to express conditionality using temporal expressions. The sentence

(50) When exposed to the air, many substances oxidize.

is used primarily to express a conditional, not a temporal relationship, though, of course, it is consistent with a temporal interpretation. The reason that temporal expressions can be used to express conditional relationships is that the most important individuating characteristic of events, the entities quantified over in conditional sentences according to Lycan and me, is the date of those events.

The parallel between _if_-clauses and _when_-clauses is a very deep one. As Chomsky (1957) noted years ago, if two strings of consituents conjoin they are normally not only syntactic constituents, but are also constituents of the same type. Consider:

(51) *John was awakened by John and by accident.

(52) *John knows that I ate an orange and what I know.

If Chomsky is correct, as the preponderance of evidence over the years would suggest, then we must assume that _if_-clauses and _when_-clauses are constituents of the same type and that _if_ and _when_ are as well.[18]

(53) I will consider leaving if I'm asked to and when I'm asked to.

(54) I will consider leaving if and when I'm asked to do so.

Greg Stump has suggested to me that _if and when_ may be an idiom, and thus that data such as these may mean very little. Against this, I would say five things. First, conjunctions of _if_-clauses and _when_-clauses are not themselves idioms, which is important since sentences containing _if and when_ merely carry the conjuction reduction a bit further. Second, *if and when* and *when and if* both occur, which is to say that the structure isn't frozen. Third, conjunction facts such as these extend to other not dissimilar phenomena (e.g. *When and where did he leave*?) which are surely not idioms. Fourth, the meaning of *if and when* is compositional, which is uncharacteristic of idioms, which is to say that each word makes a contribution to the meaning of the phrase and to the sentence as a whole. Fifth, other temporal and conditional adverbs conjoin. Note, for instance, that *until*-clauses and *unless*-clauses conjoin and *until* and *unless* conjoin.[19]

(55) I won't leave unless you leave and until you leave.

(56) I won't leave unless and until you leave.

There is clearly much too much that is *systematic* about these conjunction facts to support Stump's suggestion that we are dealing with idioms.

In light of the above, it is tempting to suggest that adverbial *if*-clauses, like adverbial *when*-clauses, are instances of the same construction and that *if* is in the same lexical class with *when*. In Geis (1970a), I argued that adverbial *when*-clauses are a species of relative clauses, and that *when* is a relative proadverb.[20] This argumentation is accepted by Larson (mss), who, working in a transformational framework, provides a somewhat different formalization.[21]

Of the various observations in Geis (1970a), the most important were that *when*-clauses can be structurally ambiguous, that *when*-clases are islands, and that *when* is a constituent of the clauses it introduces. As we shall see, *if*-clauses and *when*-clauses are similar, but not identical, in regard to these three properties.

Let me begin with the worst fact. Compare (57) and (58).

(57) I will leave when you say you'll go.

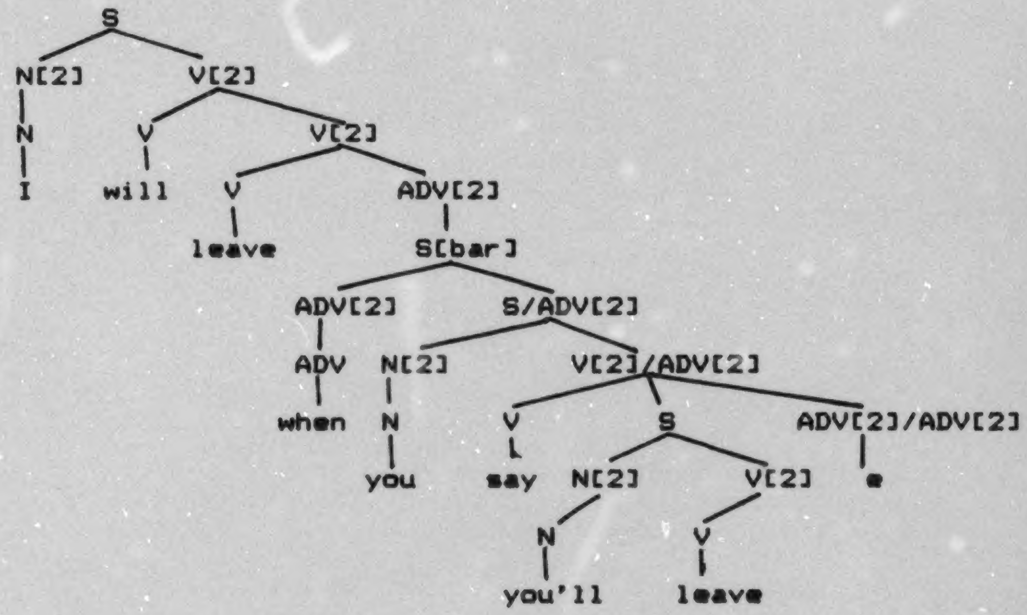
(58) I will leave at the time when you say you'll go.

Sentences (57) and (58) are ambiguous between a reading in which the speaker promises to leave when the hearer performs a speech act and one in which the speaker promises to time his departure with that of the hearer. In order to account for this syntactically, we must suppose that (57) and (58) are assigned two syntactic analyses which, in one way or another, say that *when* participates in two dependency relationships in these sentences.[22] Larson (mss) takes the transformational line of Geis (1970a) and claims that *when* is extracted from the main clause of *you say you'll leave* on one derivation and from the subordinate clause on the other. I shall take the monostratal line of GPSG and say that *when*-clauses are special cases of the construction type

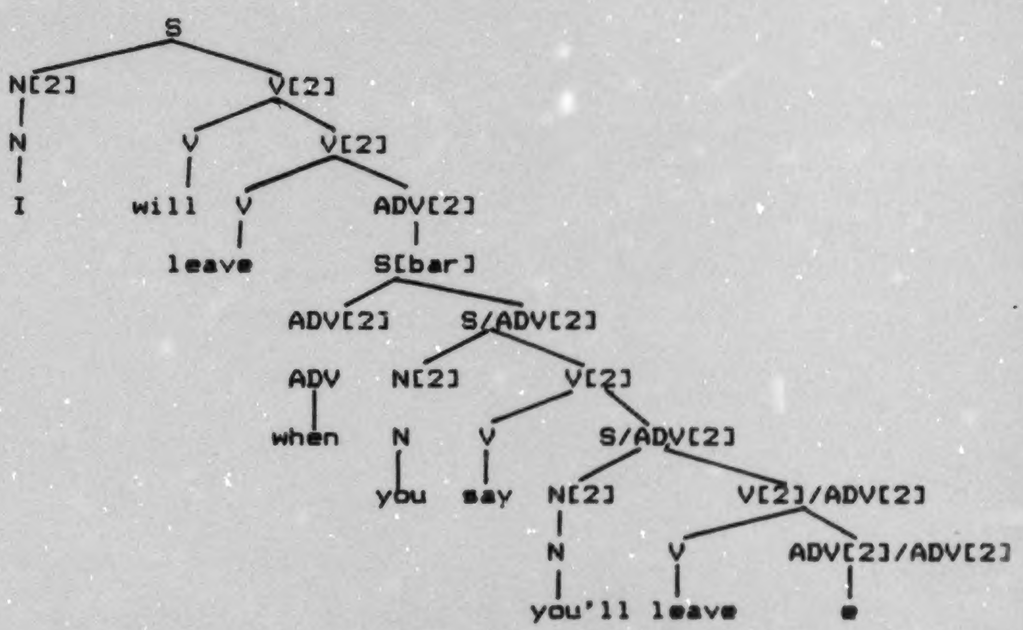
(59) S --------> α S/α (*6)

On this view (57) has a structure like (60) on one interpretation and (61) on the other.

(60)


S
N[2]
V[2]
N
V
V[2]
I
will
V
ADV[2]
leave
S[bar]
ADV[2]
S/ADV[2]
ADV
N[2]
V[2]/ADV[2]
when
N
V
S
ADV[2]/ADV[2]
you
say
N[2]
V[2]
e
N
V
you'll
leave

(61)



Contra the claim that (if)-clauses are structurally similar to *when*-clauses, that is that *if*-clauses are a species of relative clause, *if*-clauses are not ambiguous.[23] Consider (62) and (63):

(61) I will leave if you say you'll leave.

(62) I will leave in any circumstance in which you say you'll leave.

Although (62) is a bit long, and complex,[24] I believe that it is ambiguous in just the way that (58) is. But (61) is not ambiguous in the way that (57) is. Despite the otherwise overwhelming evidence supporting the thesis that *if*-clauses and *when*-clauses are grammatical sibilings this failure of the relative clause analysis cannot be disregarded. However, it is possible to attach too much significance to the nonambiguity of conditional clauses. In Geis (1970a), I presented evidence that *while*-clauses are themselves covert relative clauses despite the fact that they also are not ambiguous. Sentence (63) does not seem to have a reading in which *while* links the two occurrences of *study*.[25]

(63) I studied while Mary believed I should be studying.

However, I find sentence (64) to be quite acceptable, from which it follows that extraction is to some (albeit small) degree possible.

(64) I studied while I was supposed to.

What we are dealing with here is degrees of extractability.

Though most theoreticians resolutely ignore such messy phenomena, they exist and may say more about the nature of language than do neater, theoretically more compliant facts.[28] Moreover, there exists overwhelming semantic evidence (see Geis 1973 and Lycan 1984) that conditional clauses make covert reference to events, which is to say that sentences like (65) and (66) have essentially the same interpretations.

(65) I will leave if you leave.

(66) I will leave in any circumstance in which you leave.

The most natural syntax for a construction that connects clauses semantically via quantification over events is to say that the clauses are connected syntactically in the way relativization connects clauses. Moreover, not only are _if_-clauses unambiguous in the desired way, but so also are clauses appended to _in the event that_, which also need an analysis. Compare (67) with (61) and (62).

(67) I will leave in the event that you say

you'll leave.

In previous work (Geis 1973), I related a sentence like (61) to (67), largely because neither is ambiguous and assumed that clauses subordinated to _in the event that_ are like (nonrelative) clauses subordinated to _the fact that_, i. e. are noun complements. However, I do not believe that it is possible to give a coherent semantics for constructions like (67) given this sort of syntactic analysis. I would argue that one should adopt Lycan's semantics for (67), no less than (61), i. e. take the line that clauses embedded as adjuncts to _in the event that_ are themselves relative clauses. But if this is the right move, then the failure of _if_-clauses to mirror _when_-clauses in regard to the question of ambiguity is not fatal. Nevertheless, it must be dealt with, of course, as I shall do shortly.

The second fact supporting the thesis that _when_-clauses are relative clauses is that elements cannot be extracted out of _when_-clauses. Consider

(68) a. *Who did the boy leave town when Mary kissed <>.

b. *Who did the boy leave town at the time when

Mary kissed <>.

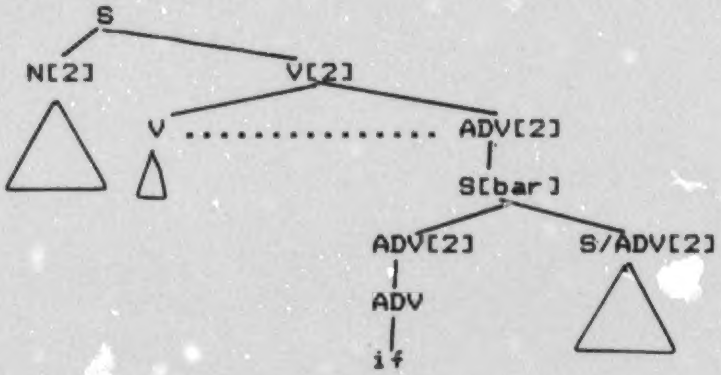
At present, there are a number of ways this sort of fact can be accounted for. I derived sentences like (68a) from sentences like (68b) in Geis (1970a) and appealed to Ross' (1967) Complex Noun Phrase Constraint to account for (68a). Though not accepting the deletion analysis, Larson (mss) gives the more recent analog of my treatment by appealing to Chomsky's (1973) Subjacency

Constraint. Within GPSG (see Gazdar (1982) for details), one would normally invoke the principle that no constituent can be assigned two slashed categories, as would be required for (68a), one for who and another for when. Not surprisingly, nothing can be extracted from if-clauses, as can be seen from

(69) *Who will the boy kiss Mary if Joe kisses <>.

In order to account for this, conditional sentences must be assigned a structure like (70) in order to invoke the condition ruling out dual slashes.

(70)



However, as we shall shortly see, this option is not available to us.

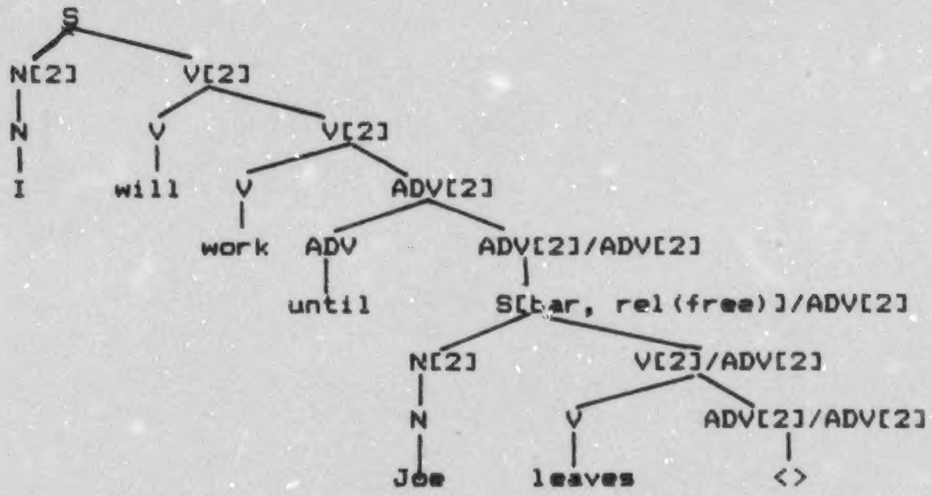
The third argument in favor of the relative clause treatment of when-clauses is that when is a constituent of the clause it introduces, after the manner of a relative proadverb. Compare (71) and (72).

(71) I will work until Joe leaves and Harry will work until then/that too.

(72) *I will leave when Joe leaves and Harry will leave when then/that too.

Manifestly (see Geis 1970a for details) the correct treatment of (71) and (72) is to say that when, but not until, is a consituent of the clauses it introduces, as in trees (60) and (61). On the other hand, the left conjunct of a sentence like (71) should be assigned a structure like (73).

(73)



Clearly, *if* functions just like *when*, for (74) is ungrammatical on the intended interpretation.

(74) *I will leave if Joe leaves and Harry will leave

if then/that too.

This fact, considered in the light of the fact that *then* can replace *if*-constructions as a whole, represents compelling evidence that *if* is a constituent of the clauses it introduces, which is to say that it is a relative proadverb. Given that *unless* conjoins with *until*, I believe that we would want to assign a tree like (73) to sentences containing *unless*, where *unless* occurs in place of *until*.

As we have seen, two of the three arguments for the relative clause treatment of *when*-clauses carry over to *if*-clauses. However, the most compelling argument derives from the fact that *then* can occur in initial position in the (grammatically) main clause of a conditional sentence if there is a sentence-initial *if*-clause. This is, I submit, a fact of over-riding syntactic importance, and when properly interpreted provides an explanation for the nonambiguity of conditional clauses. In my speech, this is a virtually unprecedented construction, but it is quite like the corelative construction that has largely departed the language. I find the following sentences to be increasingly grammatical as one moves down the list.

(75) a. ***Who steals my purse, him I won't like.

b. **Where he goes, there I'll go.

c. *When he leaves, then I'll leave.

d. If he leaves, then I'll leave.

Now consider (76).

(76) a. **Who steals my purse, I won't like.

b. *Where he goes, I'll go.

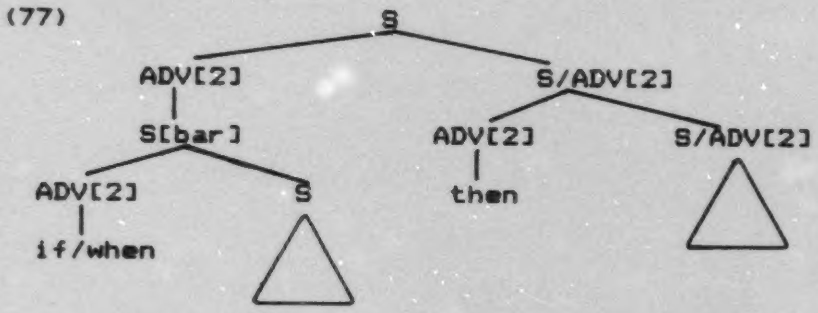
c. When he leaves, I'll leave.

d. If he leaves, I'll leave.

I suggest that the correlative construction is going out of the language, with (75d) being its remaining trace.

I would argue that sentences[27] like (75c) and (75d) have the following syntactic structure:

(77)



Quite surprisingly, correlative when-clauses, unlike conventional (i. e. post-verbal) adverbial when-clauses, are unambiguous. Although correlatives are not fully acceptable to me, I feel reasonably confident in the judgement that the when-clauses of (78) and (79), unlike that of (80), are unambiguous.

(78) ?When you say you'll phone, then I'll leave.

(79) When you say you'll phone, I'll leave.

(80) I'll leave when you say you'll phone.

Perhaps more persuasive will be (81)-(83), which demonstrate that there is no extraction reading for preposed when-clauses of either sort.

(81) *When you said you'll phone, then I'll leave.

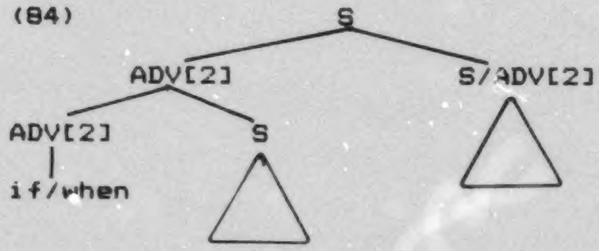
(82) *When you said you'll phone, I'll leave.

(83) I'll leave when you said you'll phone.

Thus, if we associate conditional sentences with sentences containing correlative clauses, the relative clause analysis of conditional sentences escapes unscathed from my nemesis counterexample.

Then nonambiguity of coorelative *then*-clauses is accounted for by the above analysis. Note that the S which is immediately dominated by the S[bar] of (77) is not slashed, which can be exploited by the semantics to force a "highest" clause analysis of modification by *when* and *if*. Neither can "reach" more deeply into this S than to the highest verb. I propose that preposed *when*-clauses and *if*-clauses such as those in (76c) and (76d) that occur with main clauses not preceded by *then* will also occur in structures like (84).

(84)



In my view, the *then* that fronts main clauses when *if*-clauses are preposed, is the same word that occurs in discourses such as the following:

(85) A: I'll leave at noon.

B: Then I'll leave at three.

An interesting consequence of this is, since *then* clearly has an interpretation in (85), is that it should also contribute to the meaning of a sentence like (75d). Interestingly, just as my analysis predicts, Davis (mss) has pointed out that pairs like (86) and (87) do not have the same interpretation.

(86) If you open the refrigerator, it won't explode.

(87) If you open the refrigerator, then it won't explode.

As Davis notes, (86) is true of ordinary refrigerators, while (87) is true only of refrigerators rigged to explode unless opened. I see this as especially strong evidence of the semantic benefits of the present syntactic analysis.

Conclusion

I have provided quite a number of arguments in support of an adverbial analysis of conditional clauses in general and of an adverbial relative clause analysis in particular. In a sense, this comes down to arguing that *if* is a constituent of the clauses it introduces. On the other hand, *unless* is treated as constituent of main clauses, like its morphologically similar temporal cousin *until*. These may seem to be rather small potatoes to those not versed in monostratal syntax, but this is an impression worth correcting. Whatever one's theory of syntax, one must get the surface structures of sentences right to get much of anything else right. Even transformational theories are houses of cards built on surface structure piles even though they may seem to be grander.

Perhaps the most important feature of the present analysis is that it brings conditional clauses into line with other types of adverbial clauses. In this connection, I should, perhaps, point out that *although*-clauses, which I have never had much to say about, fall out rather nicely in terms of this analysis. What they are is simply the factive counterpart of *if*-clauses, which are modified by *even*. I propose to assign them essentially the same analysis as *unless*-clauses get. Lycan's semantics can easily be expanded to include them. Just how *because*-clauses fit into the program is not as clear, though it would be surprising if they were not also to involve quantification over events.

This work was once regarded as quite abstract, for it involved postulating antecedents for adverbial relative clauses introduced by *when*, *while*, and *where* and antecedents and relative adverbs for clauses introduced by *before*, *after*, *until*, and *since*. Interestingly, the most essential syntactic features of this analysis are accomodated quite easily within the monostratal framework, GPSG, resulting in a description which is no less insightful syntactically than the transformational treatment. As a result, I believe the analysis must be all the more persuasive, since it is syntactically more conservative.

---

1. Interestingly, traditional grammarians, who do not seem to have been much influenced by logicians, did not single out conditional clauses as being of radically greater importance than other types of adverb clauses.

2. This preference of logicians, who are linguistically naive in their own way, is itself of interest, as is the fact that they virtually always cite conditional sentences with the *if*-clause preposed. See the example sentences cited in Harper, Stalnaker, and Pearce (1981) for confirmation of these points.

3. See Clark and Clark (1977) for an interesting discussion of this point.

4. Though our research is done quite separately, Lycan and I are collaborating on the development of a general theory of conditional sentences. This effort emerged out of a course Lycan

and I once jointly taught at The Ohio State University. We were examining Geis (1973), a paper in which I provided a syntactic analysis of sentences like (1d) and (2)-(6) and argued for the view that a correct semantic analysis of conditional sentences must employ quantification over events or circumstances, a semantical point of view that has come to be quite fashionable (Barwise and Perry 1983). This semantic analysis led Lycan to give the essentials of his very much more sophisticated semantic treatment, which in turn inspired me to redo significantly my syntactic treatment.

5. The reason I say this equation is question begging is that it was believed correct at that time to use semantic evidence (e.g. coocurrence data) in determining the Deep Structure of a sentence. Obviously, use of semantic data in the study of Deep Structures of sentences will have as an inevitable result that Deep Structures be Logical Forms.

6. I am not recanting the views expressed in Geis (1984) and Fox and Geis (1984) about the limitations of people's logical capacities. But the view that people do not control the validity-invalidity distinction does not require us to believe that people are not able to recognize (at least roughly) some synonymy relationships.

7. Bresnan and Grimshaw (1978) made no reference to Geis (1970a), which is perhaps due to the fact that MIT disertations are hard to come by even for those who teach at MIT.

8. I shall show below that the minimalist syntactic theory, Generalized Phrase Structure Grammar (GPSG), proposed originally by Gazdar (1981), and pursued in Gazdar (1982), Gazdar and Pullum (1982), and other papers provides sufficient descriptive apparatus to state this analysis, despite its admirably restrictive character.

9. This analysis, taken as an analysis of the meaning of English if...,then..., has very little to recommend it. The connective if...,then... is, of course, not truth functional. This was shown in Geis (1973), is argued by Gazdar (1979) in a more general way, and is further argued by Lycan (1984).

10. Interestingly, Jespersen (1961: V.4.344f), who recognized that many of the so-called "subordinating conjunctions" (e.g. the connectives of (1) above) were morphologically similar to such things as relative pronouns and prepositions, called if a "conjunction proper." Whether or not he meant to be advocating that if is therefore grammatically just like and and or in syntax is not clear.

11. Though I know of no one who has seriously proposed that the clauses that make up conditional sentences are coordinate in character, it, is nevertheless, not a straw man position. In her doctoral thesis, Heinamaki (1974) proposed that the temporal

connectives, _when_, _before_, and _until_, etc. are coordinating conjuncitons, and this is a great deal less plausible than that _if_ is a coordinating conjunction. Her arguments against Geis(1970a) are unimpressive, to say the least.

12. I am indebted to Arnold Zwicky for this observation.

13. These very important examples are due to Lakoff (1972).

14. Reference to "($n)," where "n" is a numeral, is to a rule number in the Fragment following the text.

15. Arguments of a linguistic character (i.e. arguments that are not wholly semantic) that conditional adverbials generally and _if_-clauses in particular make covert reference to events are given in Geis (1973). The Lycan (1984) paper contains a rather more sophisticated version of this analysis, with additional motivation.

16. The analysis I give of the _syntax_ of _if_-clauses can be extended to nominal occurrences of them. See rules ($5) and ($7). I treat indirect questions as [free] [int]errogatives (= [int(free)]). The only conditional [cond] proadverb [pro] that can occur in free interrogatives is unmodified _if_.

17. These names reflect the transformationalist idiom within which they were first discussed. Abandonment of this paradigm does not, of course, require that we abandon all of what can be learned from data once believed to support it.

18. Arnold Zwicky has pointed out to me that one can also conjoin _when_ and _before_ despite the fact that the former is a relative proadverb and the latter is a preposition:

(i) I will leave when or before you leave.

Because of this, I propose to treat prepositions as adverbs. In the Fragment additional motivation is given.

19. I take this as evidence that _unless_ and _until_ are in the same lexical class, which is the treatment of the Fragment.

20. To those who would object to the view that _if_ is a relative proadverb on morphophonological grounds, I would say two things. First, _how_ and _who_ differ phonetically from _what_, _when_, _where_, and _why_, but this does not stop us from saying that they, like the others, are interrogative pronouns. Second, in hosts of languages, the word used to signal conditionality is homophonous with the word used to signal "simultaneity" (and in English, as noted above, temporal words are sometimes used to signal conditional meanings.)

21. Larson's work does not include conditional sentences, so I do not know what his stand on the issues just raised would be.

22. See Stump (1981) for a semantic account of these facts.

23. When lecturing to an introductory class on English syntax taught by Edward Klima in 1964, I proposed the relative clause analysis of conditional sentences, noting this counterexample. I decided later that this sufficed to wreck the analysis. It was only on seeing Lycan's impressive reformalization of my sketchy semantics for conditionals, that I returned to this analysis. This semantic treatment clearly wants a relative clause syntax.

24. Lycan takes a paraphrase like (62) to be especially perspicuous in regard to the meanings of conditional sentences. I agree with him, and we are working toward a book-length treatment of conditionals that reconciles his intuitions with the syntactic analysis presented here.

25. See Stump (1981) and Larson (mss) for alternative interpretations of these facts.

26. As Sapir (1921) noted, "all grammars leak," and theories must be devised in which leaks are intrinsic features of grammatical descriptions rather than the embarrassments they usually are.

27. The slash category on the sister to the mother of *then* is not introduced by the rule that gives us *then*, but by Adverb Preposing ($9), the rule that positions the *when*-clause in initial position.

A Fragment

In this section, I provide an explicit chracterization of the syntax of adverbial clauses generally and conditional clauses in particular. In the process, I more fully develop the relationship between *if*-clauses and other types of conditional clauses, as well as other types of adverbial clauses, providing in the process a sketch of the motivation for the details of the analysis. I assume (a bit loosely) the framework of Gazdar and Pullum (1982), and Gazdar's (1982) treatment of relative clauses and of free relatives of the sort Bresnan and Grimshaw (1978) were concerned with.

I. Phrase Structure Rules

($1) ADV[2] ----> ADV N[2]

a. at noon, in the garden, etc.

I am treating prepositions as adverbs because some can stand alone as apparent adverbs *I haven't done that before* and can be thought of as intransitive adverbs. Those that require objects can be thought of as transitive adverbs. This approach to prepositions goes back to Jespersen (1961 II.1.15).

($2) N[2] ----> N[2] S[bar, +rel]

a. the place where Joe lives. (with $6)

b. at the time at which Joe left. (with $1 & $6)

In Gazdar and Pullum (1982) a given feature is sometimes treated as binary and sometimes treated as having other features as values. I shall exploit this by taking [rel(free)] to entail [+rel]. Though a bit equivocal, this view of features is clearly a coherent one. I shall treat [rel(free)] as the marked option for [rel]. So relatives with heads are unmarked relatives and those without heads are marked.

($3) X[2] ----> ADV[+quant] X[2]

a. Only John, only on Tuesday, even on Tuesday

This rule allows for the quantificational adverbs *only* and *even*, which I am treating as adverbials that can only modify phrasal categories (X[2]). One of the values of [+quant] is [+neg] and the other [-neg], features that play a role in triggering inversion, as will be shown below.

($4) ADV[2] ----> ADV ADV[2]

a. *up* at the barn, *until* then

b. only *until* then (with $3)

This rule allows for *adverbial* objects for certain prepositions, most of which can also occur with noun phrase objects. See Geis (1970a, 1970b) for relevant argumentation and the lexicon below for lexical details.

($5) X[2] ----> S[bar, free]

COND: [+pre] ∈ X[2] ⊃ [-slash] ∈ S[bar, free]

a. I will leave *when you leave*

b. I will go from *where you are* to *where he is*.

c. John lives near *where Bill lives*.

d. This is *where he went*.

e. *When you leave*, I'll leave.

f. *If you leave*, I'll leave.

g. I wonder *where he went*.

h. I wonder *if he went*.

i. I wonder *whether or not he went*.

j. I will leave *whether or not he went*.

This rule allows for clausal noun phrases and adverbial phrases, which are either relative or interrogative in character. I am treating embedded free relatives (a-f) and interrogatives (g-j) as instances of the class of "free" noun phrases and interrogatives. As I am using the feature, [free] is a value of [rel] and of [int], the marked value in each case. It is tempting to treat (j) as a free interrogative adverbial clause because of its similarity to (i). The condition on this rule is to insure that preposed free relatives are not slashed, i. e. are not ambiguous.

($6) S[bar] ----> (ADV[2, +wh]) S/ADV[2]

COND: [+free] ∈ S[bar] ⊃ [+pro] ∈ ADV[2]

If the mother node has the feature [+free], then the daughter node ADV[2] has the feature [+pro], a feature I use to force a monolexical pronoun for free relatives. It does not correctly get ($5h), for *whether or not* is obviously not monomorphemic, though, of course, *whether* is.

a. John lives where Joe is working. (with $5)

b. John studied until Joe left. (with $4 & $5)

c. John lives at the place where Joe lives. (with $1 and $2).

d. John lives at the place at which Joe lives. (with $1 and $2).

e. I left by the time he arrived. (with $1 and $2)

f. I will leave in any circumstance in which you leave. (with $1 and $2)

As stated this rule gets all sorts of relative clauses, including ordinary relatives with heads (c, d, e, f) and those without (a, b), which have, as "complementizers" a monolexical pronoun or proadverb (a, c), prepositional phrase (d, f), or nothing at all (b, e). This rule gets only "true" relative conditional clauses, like ($6f). To get _if_-clauses or _in the event that S_ constructions with this rule would incorrectly predict that they can be ambiguous. See the next rule.

($7) S[bar, +cond] ----> (ADV[2, +pro]) S

a. I will leave if you leave.

b. I will leave unless you leave. (with $4 and $5)

c. I won't leave unless'f you ask me to. (with $4 and $5, see also the lexical information on _unless_)

d. I will leave in the event that you leave. (with $1 and $2)

This rule gets us conditional clauses. Because the S node to the right of the arrow is not slashed, conditional clauses cannot be ambiguous. If the pronoun is [+wh] we get _if_; if [-wh], we get _that_. This distinction is required in order to get _in the event that S_ conditionals.

($8) S/ADV[2, +cond, -wh] ----> ADV[2, +pro, -wh] S

COND: [+neg] ∈ ADV[2] ⊃ [+inv] ∈ S

a. If you leave, then I'll leave.

b. If you leave, only then will I leave.

This is is the rule that gets _then_ into the main clauses of sentences with preposed _if_-clauses, which is the last remaining instance of the correlative construction most dialects. The rule is therefore ad hoc in the desired way, a synchronically explanatory theory of *if* being a theoretical pipedream. We must build into this rule the provision that if *then* is modified by *only*, its sister S must be marked as undergoing inversion. As I see it, the COND of this rule is a condition on any rule introducing ADV[2] and S as sisters, including the next rule.

($9) S ----> ADV[2, +pre] S/ADV[2]

a. At noon, John left.

b. If you leave, I'll leave.

c. Only if you leave will I leave.

d. ?Only if you leave then will I leave.

This is adverb preposing, of course. It is subject to the condition on the previous rule. If we wish to block ($9d), we will need to say that if ADV[2] is [αmod], then S/ADV[2] is [-αmod], where [quant] is a value of [+mod]. The feature [+pre] ([+pre] = 'preposed') is there to guarantee that preposed free relatives are not ambiguous--see Rule ($5).

II. Lexical entries

A. at, on, in, up, until, $unless = +Rule ($1)

B. only, even = +Rule ($3)

C. up, until, unless, although = +Rule ($4)

D. near, in front of = +Rule ($4)

COND: [+prep] ICW [+rel] ⊃ [+adv, +wh] ∈ [+rel]

The stipulation--for place prepositions, but not time prepositions--is that if *near* and *in front of* occur in construction with (ICW) a relative clause, the clause must have an overt relative proadverb. I use the notion "in construction with" here for perspicuity, and do not mean to be making the claim that this notion is required.

E. until, since, before, after, unless

COND: [+prep] ICW [+rel] ⊃ [-wh] ∈ [+rel]

The condition--for time and conditional prepositions, but not place prepositions--guarantees that relative clauses introduced by these words will not have an overt relative proadverb. For those who can say *I won't leave unless'f you ask me to*, as I can, *unless* is not in this list. I know of no analysis of conditionals that can cope at all with this datum.

F. if, when, where = [adv(pro(wh))]

G. then = [adv(pro(-wh))]

References

Baker, G. P. and Hacker, P. M. S. 1984. Language, sense and nonsense: A critical investigation into modern theories of language. Oxford: Basil Blackwell.

Barwise, J. and Perry, J. 1983. Situations and attitudes. Cambridge, MA.: MIT Press.

Bresnan, J. W. and Grimshaw, J. 1978. The syntax of free relatives in English. Linguistic Inquiry, 9.331-391.

Chomsky, N. 1973. Conditions on transformations. In S. R. Anderson and P. Kiparsky, A festshrift for Morris Halle. New York: Holt, Rinehart, and Winston.

Clark, H. H. and Clark, E. V. 1977. Psychology and language: An introduction to psycholinguistics. New York.: Harcourt, Brace, Jovanovich.

Fox, R. A. and Geis, M. L. 1984. Mental logic and sentence comprehension. Proceedings of the 1984 Regional Meeting of the Chicago Linguistic Society.

Gazdar, G. 1982 Phrase structure grammar. In P. Jacobson and G. K. Pullum, eds., The nature of syntactic representation, 131-186. Dordrecht, Holland: D. Reidel.

Gazdar, G. 1981. Unbounded dependencies and coordinate structure. Linguistic Inquiry, 12.155-184

Gazdar, G. 1979. Pragmatics: Implicature, presupposition, and logical form. New York: Academic Press.

Gazdar, G. and Pullam, G. K. 1982. Generalized phrase structure grammar: A theoretical synopsis. Bloomington, IN: Indiana University Linguistics Club.

Geis, M. L. 1984. On semantic and pragmatic competence. Georgetown University Round Table on Language and Linguistics. Washington, D.C.: Georgetown University Press.

Geis, M. L. 1970 Adverbial subordinate clauses in English. Unpublished doctoral thesis. Cambridge, MA: MIT.

Geis, M. L. 1970. Time prepositions as underlying verbs. Papers from the sixth regional meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society.

Geis, M. L. 1973. If and unless. In B. B. Kachru, et al, eds., Issues in lingustics: Papers in honor of Henry and Rene Kahane. Urbana, IL. University of Illinois.

Harman, G. 1972. Logical Form. Foundations of Language, 9.36-65.

Harper, W. L., Stalnaker, R., and Pearce, G. 1981. Ifs. Dordrecht, Holland: D. Reidel.

Heinamaki, O. 1974. Semantics of English Temporal Connectives. Unpublished doctoral thesis. Austin, TX: University of Texas.

Jespersen, O. 1956. A modern english grammar on historical principles. London: George Allen and Unwin.

Lakoff, G. 1972. Linguistics and natural logic. In D. Davidson and G. Harman, eds., Semantics of natural language, 545-665. Dordrecht, Holland: D. Reidel.

Larson, R. mss. Extraction and doubvle selection in PP.

Lycan, W. G. 1984. A syntactically motivated theory of conditionals. Midwest Studies in Philosophy.

Sapir, E. 1921. Language. New York: Harcourt, Brace, & World.

Stump, G. 1981. The formal semantics and pragmatics of adjuncts and absolutes in English. Unpublished doctoral dissertation. Columbus, OH: The Ohio State University.

Talmy, L. 1976. Semantic causative types. Syntax and semantics, (V. 6): The grammar of causative constructions. New York: Academic Press.

OSU WPL 31 (1985) 160-194

German adjective agreement in GPSG*

Arnold M. Zwicky
The Ohio State University

0. Summary

Determiners and adjectives in German agree with their head nouns in case, gender, and number. In addition, all adjectives have three paradigms of inflectional forms, which are traditionally called 'declensions': strong, weak, and mixed. Which declension an adjective occurs in depends on the determiner it combines with, a phenomenon traditionally called 'agreement'. Section 1 presents the main facts about adjective agreement in German, in a fairly theory-neutral fashion.

In a rich theory of syntax (like classical transformational grammar) it would be easy to write rules for German which have the right effects. My purpose here is to explore how to describe German adjective agreement in a distinctly lean theory of syntax, namely generalized phrase structure grammar (GPSG). Section 2 enumerates the principal features of GPSG, paying special attention to those that might figure in accounts of agreement.

Sections 3 and 4 attack the problem of the three adjective declensions. In section 3, several functional accounts of the distribution of forms are subjected to scrutiny and found wanting. In section 4, GPSG descriptions treating the phenomena as subcategorization are shown to be unsuitable, and those treating them as agreement are shown to be unavailable. The appropriate GPSG analysis involves _government_ rather than agreement, a conclusion that leads to some general comments on the description of government in GPSG.

1. The facts

I begin with a reasonably precise, though unformalized, account of the relevant German facts.

The language has three grammatical genders (masculine, neuter, and feminine) and two grammatical numbers (singular and plural). Only four of the six combinations of gender and number are ever morphologically distinguished: MASC-SG, NEUT-SG, FEM-SG, PLURAL.

There are four grammatical cases: nominative (NOM), accusative (ACC), genitive (GEN), dative (DAT). These combine with the four gender/number possibilities to yield a paradigm with sixteen potentially distinct NP forms in it.

I now turn to a summary of the adjective declensions (section 1.1), the corresponding classes of determiners (1.2), and the declensions to which the determiners themselves belong (1.3), with a summary of these matters in section 1.4. Declensions of nouns are treated in section 1.5, where it is pointed out that the 'declensions' of nouns and determiners are lexical categorizations, whereas the 'declensions' of adjectives in German are imposed by constituents with which the adjective is in construction. Section 1.6 looks briefly again at the grammatical categories of gender, number, and case, in comparison to the declension categories. And section 1.7 summarizes the whole business.

1.1. *The adjective declensions*

German adjectives occur in three paradigms of forms:

--'strong' forms, which occur with a zero determiner or with an invariable determiner like *zwei* 'two';

--'weak' forms (with massive levelling of the distinctions marked in the strong forms), which occur with *der* 'the', *dieser* 'this', *jener* 'that', *welcher* 'which', and *solcher* 'such';

--'mixed' forms (with some endings from each of the two other sets), which occur with *ein* 'one, a', *kein* 'no', and the possessive pronouns (*mein* 'my', *unser* 'our', and so on).

The strong forms can be seen in *gut-er Mann* 'good man' and *gut-e Maenn-er* 'good men' (here I have indicated morpheme breaks by a hyphen); the weak forms in *dies-er gut-e Mann* 'this good man' and *dies-e gut-en Maenn-er* 'these good men'; the mixed forms in *kein gut-er Mann* 'no good man' and *kein-e gut-en Maenn-er* 'no good men'.

The endings for the three sets are shown in Tables 1-3. Six of the sixteen case/gender/number combinations have the same endings in all three sets; these six are underlined in the tables. Notice that the mixed declension is indeed an amalgam of endings from the strong and weak declensions, though with the weak declension predominating: of the ten endings that differ in the strong and weak sets, the mixed set takes seven from the weak and three from the strong. With some justification, we might then consider the mixed declension as a special subtype of the weak declension. In what follows, I will call the strong declension 'Declension S', the weak declension 'Declension W', and the mixed declension 'Declension W-MX'.

CASE	MASC-SG	NEUT-SG	FEM-SG	PLURAL
NOM	-er	-es	-e	-e
ACC	-en	-es	-e	-e
GEN	-en	-en	-er	-er
DAT	-em	-em	-er	-en

Table 1. Strong adjective endings.

CASE	MASC-SG	NEUT-SG	FEM-SG	PLURAL
NOM	-e	-e	-e	-en
ACC	-en	-e	-e	-en
GEN	-en	-en	-en	-en
DAT	-en	-en	-en	-en

Table 2. Weak adjective endings.

CASE	MASC-SG	NEUT-SG	FEM-SG	PLURAL
NOM	-er	-es	-e	-en
ACC	-en	-es	-e	-en
GEN	-en	-en	-en	-en
DAT	-en	-en	-en	-en

Table 3. Mixed adjective endings.

1.2. The determiner classses

Several remarks should be made about this array of facts. First, the grouping of factors conditioning strong vs. weak vs. mixed adjective declension is not semantic, at least not on any account I can imagine. In particular, the determiners conditioning strong declension include both definites (like *zwei*) and indefinites (like the zero determiner and the exclamatory indeclinable determiner *welch*); the determiners conditioning weak declension also include both definites (like *der*) and indefinites (like *welcher*); and the determiners conditioning mixed declension also include both definites (like *mein*) and indefinites (like *ein*). That is, it appears that the division of determiners into three classes is a grammatical, and not a semantic, classification. The three classes might as well be named 'Class I', 'Class II', and 'Class III'--and, indeed, in what follows I will use this nomenclature.

1.3. The determiner declensions

Next, two of the three classes of determiners are declinable, and for each class there is a single declension type. The endings for Class II determiners are shown in Table 4, those for Class III determiners in Table 5. Notice that the endings for Class II determiners are almost identical to those conditioned by Class I determiners--only the masculine and neuter genitive singular endings (both -es)[1] differ--and that the endings for Class III determiners are closer to the adjective declension conditioned by Class I determiners than to either of the others (the Class III determiners share eleven of their sixteen endings with adjectives conditioned by Class I determiners, and only four endings with adjectives conditioned by Class II or Class III determiners).

1.4. Summary of the data so far

To sharpen, and abbreviate, the observations of the previous paragraph: Class I determiners are indeclinable; Class II determiners belong to a subtype, call it 'Declension S-ES', of Declension S (with a special ending -es in the masculine and neuter genitive singular); Class III determiners belong to a subtype, call it 'Declension S-ES-Z', of Declension S-ES (with zero endings in the masculine and neuter nominative singular and the neuter accusative singular, as well as the special ending -es in the masculine and neuter genitive singular). In other words, though determiners of Classes II and III condition adjectives of Declensions W and W-MX, respectively, the determiners themselves belong to (subtypes of) Declension S.

CASE	MASC-SG	NEUT-SG	FEM-SG	PLURAL
NOM	-er	-es	-e	-e
ACC	-en	-es	-e	-e
GEN	-es	-es	-er	-er
DAT	-em	-em	-er	-en

Table 4. Declension of Class II determiners.[2]

CASE	MASC-SG	NEUT-SG	FEM-SG	PLURAL
NOM	zero	zero	-e	-e
ACC	-en	zero	-e	-e
GEN	-es	-es	-er	-er
DAT	-em	-em	-er	-en

Table 5. Declension of Class III determiners.

1.5. *The noun declensions*

A final complexity is that nouns also have several declensional patterns, also traditionally described in terms of 'strong', 'weak', and 'mixed' types (as in Curme 1960: 70-94). The strong noun declension is summarized in Table 6, the weak in Table 7, and the mixed in Table 8.

The strong declension has one of three plural markers, indicated by PL in Table 6: -*e*, with a zero allomorph, as in *Arm-e* 'arms' and *Engel* 'angels'; -*e*, also with a zero allomorph, accompanied by umlaut in the base, as in *Soehn-e* 'sons' and *Brueder* 'brothers'; and -*er* accompanied by umlaut in the base, as in ~~Buech-er~~ 'books'.[3]
The weak declension (which contains no neuter nouns) has -*en* throughout the plural.

The mixed noun declension can be seen simply as a type of strong declension with -*en* as the plural marker, and I will do so here. One further type of noun declension, used especially for foreign borrowings like *der Domino* 'domino'--with -*s* in the genitive singular of masculine and neuter nouns and throughout the plural, and zero endings otherwise--is also clearly a subtype of the strong declension. Consequently, I opt for an analysis with only two declension classes for nouns, Declension S (strong) and Declension W (weak).

What is important here is that nouns, like determiners but unlike adjectives, are individually (and essentially arbitrarily) assigned to particular declension classes. *Mann* 'man' belongs to the strong declension (of the subtype with umlauted -*er* plurals), but *Knabe* 'boy' belongs to the weak declension. Declension class is a lexical property of particular nouns and determiners; adjectives, however, belong to no declension class lexically, but are assigned to a class by virtue of the type of determiner with which they are in construction. The declension class of the *noun* with which an adjective is in construction plays almost no role (but see Durrell 1979: 71) in determining the declension class of an adjective: *dies-er gut-e Mann* 'this good man' has a weak adjectival form occurring with a noun belonging to the strong declension; *gut-er Mann* 'good man' has a strong adjectival form with the same noun; *dies-er gut-e Knabe* 'this good boy' has a weak adjectival form occurring with a noun belonging to the weak declension; and *gut-er Knabe* 'good boy' has a strong adjectival form occurring with the same noun.

1.6. *Other grammatical categories*

Gender, like declension class, is a lexical property of particular nouns; nothing predicts that *Arm* 'arm' is masculine and *Hand* 'hand' feminine. The gender classification of both adjectives and determiners is determined by the gender of the noun with which they are in construction: *dies-er Arm* 'this arm' and *gut-er Arm* 'good arm', *dies-es Buch* 'this book' and *gut-es Buch* 'good book', *dies-e Hand* 'this hand' and *gut-e Hand* 'good hand'.

The remaining grammatical categories that play a role in German adjective inflection, number and case, are in general not lexical properties of any word class. However, a determiner, adjective, and noun in construction with one another must agree in both number and case. Case is, of course, a property of whole noun phrases, determined by the syntactic context in which they occur. I will assume that number is also a property of whole noun phrases, one that is 'freely chosen' rather than determined by context.

1.7. *Summary of the facts*

--Declension is a lexical property of nouns and determiners, but not adjectives; nouns are essentially either Declension S (strong) or Declension W (weak), and determiners either belong to a subtype of Declension S or are indeclinable;

--Determiners are lexically (and arbitrarily) assigned to Class I, Class II, or Class III;

--The declension of an adjective is determined by the class of the determiner with which it is in construction (Declension S for a determiner of Class I, Declension W for a determiner of Class II, and Declension W-MX for a determiner of Class III);

--Gender is a lexical property of nouns, but not of determiners or adjectives;

--The gender of a determiner or adjective is determined by the gender of the noun with which it is in construction;

--Case and number are assigned to noun phrases as wholes;

--And the case and number of an NP must be duplicated as properties of the determiner, adjective, and noun within that NP.

CASE	MASC-SG	NEUT-SG	FEM-SG	PLURAL
NOM	zero	zero	zero	-PL
ACC	zero	zero	zero	-PL
GEN	-(e)s	-(e)s	zero	-PL
DAT	-(e)	-(e)	zero	-PL-n

Table 6. Strong noun declension.

CASE	MASC-SG	FEM-SG	PLURAL
NOM	zero	zero	-en
ACC	-en	zero	-en
GEN	-en	zero	-en
DAT	-en	zero	-en

Table 7. Weak noun declension.

CASE	MASC-SG	NEUT-SG	PLURAL
NOM	zero	zero	-en
ACC	zero	zero	-en
GEN	-(e)s	-(e)s	-en
DAT	-(e)	-(e)	-en

Table 8. Mixed noun declension.

2. Generalized phrase structure grammar

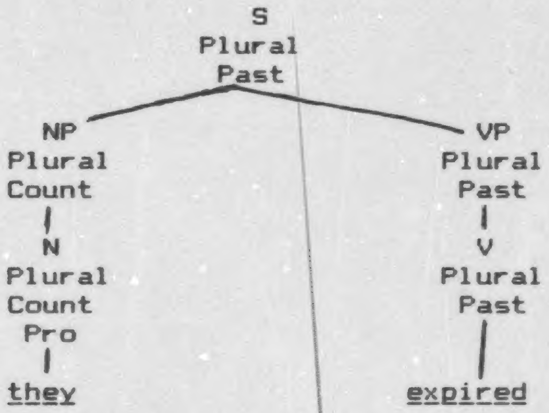
Given the above facts about the occurrence of certain inflected forms in German, my task is now to turn this relatively theory-neutral account into at least a sketch of a precise description. There are a number of theoretical frameworks in which such a description could be couched. Transformational grammar in its many varieties, for instance, easily permits rules to be stated which will require one node in a syntactic structure to bear certain features on the basis of the features borne by other nodes.

However, for my exposition I have chosen the theoretical framework of generalized phrase structure grammar (GPSG), especially as developed by Gazdar and Pullum (1982; hereafter GP), because the framework is highly constrained, both in an exact technical sense and also in a looser sense. The describable sets of strings in GPSG (as defined by GP) are all context-free languages; that is the technical sense in which GPSG is a restricted framework. Independently of its restriction to context-free languages, GPSG attempts to place universal restrictions on the sorts of syntactic rules languages can have, and consequently on the set of possible languages; that is the loose sense in which GPSG is a restricted framework. (What makes it loose is that restricting the set of grammars does not necessarily restrict the set of languages generated, as Wasow (1978) has emphasized.)

In the remainder of this section, I describe the central features of GPSG. Some of these are shared with other current syntactic theories, others are especially characteristic of GPSG. Most have some bearing on the description of German adjective agreement.

2.1. *Context-free rules*

GPSG requires that all syntactic rules be context-free. That is, every syntactic rule in a language describes a possible *branching*, of a 'mother' category into a set of 'daughter' categories, in constituent structures in that language. A full constituent structure is consistent with the grammar if all the branchings in it are described by rules for that language. To say that



is a constituent structure of English is to say that each of the three branchings in this constituent structure--Plural Past S branching into a Plural Count NP and a Plural Past VP, Plural Count NP branching into a Plural Count Pro N, and Plural Past VP branching into Plural Past V--are licensed by the syntactic rules of English, and that the lexicon of English includes _they_ as a Plural Count Pro N and _expired_ as a Plural Past V.

2.2. _Decomposition of categories_

In common with virtually all current syntactic theories derived from, or framed in response to, classical transformational grammar, GPSG decomposes categories into sets of properties. Thus, a category like NP is decomposed into two components, one indicating that it is a noun-type, or _nominal_, category, the other indicating that it is a 'two-bar', or _phrasal_, category; this decomposition can be represented by the following notation, which has the spirit of GP's proposals, while differing from it in details:

{CAT:N, BAR:2}.

In the same vein a Plural Count NP would get a representation like

{CAT:N, BAR:2, NUM:+, CNT:+}.

In such representations, a _property_ like CAT:N is actually a pairing of an _attribute_, here CAT, and a _value_, here N.

The version of GPSG given by GP treats categories as _complexes_ of properties, with internal structure. In particular, there are significant subtypes of properties within a category. GP distinguish (a) 'head' properties (they call them 'head features'), (b) 'foot' properties (they call them 'foot features'), and (c) properties that are neither head nor foot properties; within the set of head properties, they distinguish (a1) 'agreement' head properties from (a2) all other head properties; within the set of foot properties, they in effect distinguish between (b1) foot properties (like reflexivity and wh-ness) that occur in lexical entries and (b2) the special 'slash' foot property, which is used in GPSG analyses of constructions with gaps in them. These distinctions in nomenclature correspond to different sorts of conditions on the occurrence of properties in branchings, but for the moment let me simply stipulate that it is necessary to refer to two subsets of the properties within a category, and also to refer in turn to a subset of one of these.

To represent this categorial substructure, I will follow GP in treating HEAD, FOOT, AGR, and SLASH themselves as attributes, taking sets of properties--that is, categories--as values. An example will clarify the proposal. I will suppose that number and case are head

properties, whose attributes are NUM and CASE, respectively; that number and case are agreement properties; that reflexive constituents belong to a category having a foot property with the attribute REFL; and that CAT and BAR are attributes of properties that are neither head nor foot properties. Given all of these assumptions, the representation of an accusative plural reflexive NP would be
{CAT:N, BAR:2, HEAD:{AGR:{NUM:+, CASE:ACC}}, FOOT:{REFL:+}}.
Similarly, a plural clause with an NP 'hole' in it would have a representation like
{CAT:V,BAR:3,HEAD:{AGR:{NUM:+}},FOOT:{SLASH:{CAT:N,BAR:2}}}.

2.3. Metagrammars

An obvious stumbling-block lies in the path of anyone who maintains that a grammar for a language is nothing but a set of context-free rules describing possible branchings in that language: The number of such rules, in any language, is huge; and in any case merely enumerating this gigantic list utterly fails to express any generalizations about constituent structures.

GPSG's response to this objection is to generate rather than list the rules. Since each context-free rule is a description of an elementary piece of constituent structure, generalizations about constituent structures can be stated as generalizations about the set of rules, in a 'metagrammar' that describes the content of the grammar itself. The principles in this metagrammar might be of many types--some universal, some language-particular; some summarizing sets of rules in a single formula, some deriving sets of rules from a rule prototype, some predicting the existence of sets of rules on the basis of the existence of other sets--but most of these details need not concern us here. What is important is that the general program, of describing a large set of context-free rules in terms of general principles, is at least plausible.

Certain features of this program are important to us, however. These are treated in the next two subsections.

2.4. Free instantiation, implications, and defaults

One important issue for us is how to describe conditions on the co-occurrence of properties within categories. Consider, as an example, how to describe the fact that in German it is generally the case that the number of an NP can be 'freely chosen'--that is, does not depend on the number of neighboring constituents. The apparent difficulty is that essentially every time we want to state a rule introducing NP, we must state two rules, one to introduce Singular NP and one to introduce Plural NP. A generalization is being missed.

One solution to the problem is to state a prototype rule that does not mention number but merely introduces NP, and to let the values of the NUM property be 'freely instantiated'. Every such prototype rule then acts as an abbreviation for two rules, one mentioning the property NUM:- and one mentioning the property NUM:+. In general, we can suppose that

> Property values are freely instantiated, except where this would be contradicted by some other principle of the metagrammar.

The very opposite sort of situation also occurs, of course. In such cases, the range of values for some property is completely determined by other property values within the same category. For instance, in section 1.5 above we remarked in passing that there are no neuter nouns with the weak declension; that is, a noun with the weak declension is either masculine or feminine. The sort of general principle we need to state here happens to be language-particular, but what is important is that it is implicational in form: if N has the property DECL:W, then N also has the property GEND:M or GEND:F. Thus,

> The metagrammar includes principles predicting the range of values for one property on the basis of the values of other properties within the same category.

In some cases the relationship between properties within a category is not implicational (in the sense that one set of properties *requires* another), but 'nearly implicational' (in the sense that one set of properties is *usually associated* with another).

An instance of this latter relationship in German concerns the grammatical case of the direct object of a verb: There are verbs that require their direct objects to have dative case (*aehneln* 'resemble', for instance), and verbs that require their direct objects to have genitive case (*genesen* 'be delivered of, give birth to', for instance), but nearly all verbs require (or permit) their direct objects to have accusative case. We cannot say that if an NP is the direct object of a verb, then it is accusative--but we *can* say that if an NP is the direct object of a verb, then in the absence of further information we expect it to have accusative case. Accusative case is the *default* assignment of case to direct objects in German. In general, then, we want to be able to say that

> The metagrammar includes principles that assign a certain value to some property within a category in the absence of some other principle assigning a value to that property in that category.

My example happens to be specific to German, but universal default assignments are also possible. For instance, I will assume that the default value for any foot property is −; the German lexicon must provide the information that the determiner welcher has the foot property WH:+, and the English lexicon must do the same for the determiner which, but neither lexicon has to specify that dieser or unser (in German) or this or our (in English) has the foot property WH:−.

2.5. Principles governing property agreement

The reason we want to distinguish head properties from foot properties, and agreement head properties from other head properties, is that principles can be formulated that govern the way in which each type of property can occur in branchings. A significant claim made by GP is that much of the content of these principles is universal rather than language-particular; I will comment on this aspect of feature-agreement in the next section.

Given the GPSG proposal that rules describe nothing more than a mother category and its daughter categories, there can be only two types of conditions on the co-occurrence of properties between constituents: those relating the properties in the mother category and the properties in (one or more) daughter categories; and those relating the properties in two (or more) daughter categories under the same mother. GP suggest conditions of both types.

Conditions on the co-occurrence of properties could take many forms, of course. As it happens, the three conditions proposed by GP are all positive, rather than negative, and (in combination with assumptions about free instantiation, implications, and defaults) they all have the effect of requiring that certain properties agree, that is, have the same values. Two of the conditions, the Head Feature Convention and the Foot Feature Principle, govern mother-daughter property agreement; the remaining condition, the Control Agreement Principle, governs property agreement between certain pairs of sisters.*

The Head Feature Convention (HFC) ensures that the head properties in a mother category and the head properties in the daughter category that is the head of the construction are identical. Assuming that the internal structure of a German (or, for that matter, English) NP involves the branching of NP into Det and Nom, Nom into AP and N, AP into A', and A' into A, then the HFC ensures that the head properties in the following pairs of categories are identical: NP and Nom, Nom and N, AP and A', A' and A. Rather more precisely, given a rule prototype that licenses the branching of

{CAT:N, BAR:2}

into the daughters

{CAT:Det, BAR:0} {CAT:N, BAR:1},

the HFC permits the branching of

{CAT:N, BAR:2, HEAD:{AGR:{NUM:+, GEND:F}}}

into the daughters

{CAT:Det, BAR:0} {CAT:N, BAR:1, HEAD:{AGR:{NUM:+, GEND:F}}}

but does not permit the branching of

{CAT:N, BAR:2, HEAD:{AGR:{NUM:+, GEND:F}}}

into the daughters

{CAT:Det, BAR:0} {CAT:N, BAR:1, HEAD:{AGR:{NUM:-, GEND:M}}}.

Here, the HFC requires identity of the head properties in NP and Nom; similar statements can be made for Nom and N, AP and A', and A' and A.

The Control Agreement Principle (CAP) interacts with the HFC to describe grammatical agreement in languages. Given a list of what I will call agreement pairs, certain pairs of sister categories, the CAP has the effect of ensuring that the two sister categories in a pair have the same agreement head properties. The list of agreement pairs--for the moment we do not have to be concerned here with where this list comes from--includes NP and VP, Det and Nom, AP and N. Then, given the branching (just above) of

{CAT:N, BAR:2, HEAD:{AGR:{NUM:+, GEND:F}}}

into

{CAT:Det, BAR:0} {CAT:N, BAR:1, HEAD:{AGR:{NUM:+, GEND:F}}},

the CAP requires that the properties in Det fill out to

{CAT:Det, BAR:0, HEAD:{AGR:{NUM:+, GEND:F}}}.

In German, the HFC and CAP together ensure that determiners, adjectives and nouns in construction with one another have the same values for the properties of number, case, and gender. Speaking very loosely, gender markings 'originate with' the lexical item N, while number and case markings 'originate with' the NP node dominating the whole business. The HFC requires that the gender marking on N be duplicated on Nom and then on NP; the CAP requires that the gender marking on Nom be duplicated on Det; the CAP also requires that the gender marking on N be duplicated on AP; and the HFC ultimately requires that the gender marking on AP be duplicated on A. As for case and number, the HFC requires that their markings on NP be duplicated on Nom and then N, and the CAP and HFC, as before, require that these markings be reproduced ultimately on Det and A.

The third agreement principle, the Foot Feature Principle (FFP), requires that a mother category possess every foot property appearing in any one of its daughter categories. In GP's treatment, the FFP acts as a constraint on the free instantiation of foot properties, and only as such a constraint; it does not 'propagate' properties appearing in categories by virtue of rule or metarule application.

2.6. Universal aspects of property-agreement principles

GP assume that all the content of the HFC, FFP, and CAP is universal. What is potentially particular to a given language, on this view, is the list of head properties; the specification of which daughter constituent is the head of a construction; the list of foot properties; the list of agreement head properties; and the list of agreement pairs.

In fact, GP entertain two further restrictions on language-particular variation. First, they observe that X-bar syntax generally assumes some universal principle (referring to category membership and bar level) that picks out the head constituent (if there is one). Their own proposal takes a somewhat different tack, marking heads explicitly but then using the HFC to predict their category membership. In any event, it seems clear that selecting the head and assigning it category membership are not independent operations.

Second, GP propose that the list of agreement pairs be universally determined. Indeed, they propose (building on ideas in Keenan 1974) that the list can be derived from the semantic principles associated with syntactic branchings; their statement of the CAP requires that two syntactic constituents standing semantically in a 'controller'-'controllee' (roughly, argument-functor) relationship have the same agreement head properties. I will not explore this proposal here. It is sufficient to observe that on any reasonable interpretation, the CAP will require that German nouns and their accompanying adjectives and determiners all have the same agreement head properties.

If universal versions of the HFC and the CAP are to 'provide the basis for a highly effective theory of agreement' (GP, 31), then the interaction of these two principles must be the only source of systematic agreement in head properties between two categories neither of which dominates the other; in particular, the CAP must be the only source of systematic agreement in head properties between two sister categories. The GP proposal for agreement would be completely undercut if there could be language-particular (meta)rules requiring identity of properties between sister categories. There is already genuine variation from language to language as to which properties are agreement properties, including the possibility that the set of agreement properties is empty. If languages with an empty set of agreement properties could nevertheless have idiosyncratic agreement rules, then there would be no pattern of property agreement or disagreement that could not be given a description; the CAP would not constrain grammatical theory at all. We appear to need something like the following Property Agreement Restriction (PAR):

No language-particular (meta)rule requires agreement in one or more properties between two sister constituents.

One further aspect of GP's treatment of agreement needs amendment here. GP do not constrain the set of head (or foot) properties in any way, but it was surely not their intention to suggest that each language could select its own set of head properties to function in the HFC and CAP and its own set of foot properties to function in the FFP. Rather, universal grammar should permit only a finite number of such properties; indeed, universal grammar should provide finite lists of the properties available for service in any particular grammar, a Universal Head Property List (UHPL) and a Universal Foot Property List (UFPL). The attributes on the UHPL correspond to the familiar grammatical categories of person, number, gender, definiteness, case, tense, aspect, voice, mood, negation, and the like. The attributes on the UFPL include at least WH, REFL, and SLASH.

In referring to properties like NUM:+, GEND:F, and WH:-, I am insisting that the properties on the UHPL and UFPL are not mere formal counters (not just the names 'NUM:+', 'GEND:F', and so on), that they have some substance. In particular, I require that every property on the lists have semantic concomitants. I am not maintaining here that these properties are to be identified with semantic features; grammatical categories are virtually always arbitrarily distributed in the lexicon to some extent. I am maintaining that head and foot properties are never fully arbitrary and language-particular categorizations of words and phrases; if they could be, then there would be no point in having a UHPL and UFPL. Fully arbitrary and language-particular categorizations of words are indeed possible--declension classes of nouns and conjugation classes of verbs are clearly like this in some languages--but, assuming the UHPL and UFPL, lexical properties of this sort cannot be either head or foot properties and so cannot be subject to the HFC, CAP, or FFP; and, assuming the PAR, they cannot be subject to language-particular agreement (meta)rules either. These parochial properties are not subject to any sort of agreement principles.

To summarize: Parochial properties play no role in any sort of agreement relationships, and are not drawn from a substantive universal list. In contrast, agreement properties are distributed via the HFC and CAP, and since they are head properties, they must be chosen from a universal list and cannot be invented afresh for each language.

2.7. Two distinct types of 'agreement'

The CAP is designed to cover only phenomena of grammatical agreement, in a narrow sense. It provides no account of agreement between anaphoric elements and their antecedents, as when it is said that the pronoun *er* agrees with its antecedent *der Mann* in the sentence *Der Mann sagt, dass er krank ist* 'The man says that he is sick'. Anaphor-antecedent agreement in GPSG needs a different sort of account from the one the theory makes available for describing (for instance) the agreement in gender and number between the article *der* and the noun *Mann* in this example. The analysis of anaphor-antecedent agreement will be closely tied to rules of semantic interpretation, perhaps via a general principle like Lapointe's (1983: 125) Well-formedness Condition on S-structures, which says that 'If two word-level categories in a S-structure are logically connected, then they must agree on whatever non-semantic morphological features they share'.[8]

2.8. Lexical subcategorization by rule index

Just as they refrain from attempting a uniform semantic account of agreement, choosing instead to describe some facts entirely via syntactic rules and others in part by reference to semantic interpretation, so GP reject thoroughly semantic accounts for the subcategorization of lexical items with respect to the set of sister categories they can combine with. Instead, they argue that at least some subcategorization facts require a syntactic treatment.

They propose assigning each phrase structure rule an index and letting this index be represented as a property in any lexical category introduced by the rule. If, for instance, rule 6 expands NP as Det Nom, then the Det introduced by the rule will have the index 6 represented as one of its properties. And any determiner that can combine with a Nom will have the index 61 represented as one of its properties in the lexicon.

2.9. Other features of GPSG

The remaining characteristic features of GPSG do not play a central role in my discussion of German adjective agreement. I mention them here for completeness.

The first of these (already mentioned above) is the use of a foot property with the attribute SLASH to describe gap-filler dependencies, for instance the dependency between a gap within a relative clause and the relative pronoun that serves as its filler.

Next is the 'IDLP format', which allows grammars to be framed in a way that separates principles describing constituency (immediate dominance) from those describing the linear ordering of coconstituents.

Finally, there is the rule-to-rule semantics of GPSG, the assumption that a rule comprises both a syntactic part (specifying a permissible branching into constituents) and a semantic part (a function specifying a semantic interpretation for the whole construct, given as arguments the semantic interpretations of those constituents). On this assumption, two rules are distinct, and so get distinct indices, if either their syntactic parts or their semantic parts are distinct.

3. Analyzing the declensions: functional proposals

The GPSG framework of section 2 permits a satisfactory description of many details about the forms that German prenominal adjectives take. In particular, agreement with respect to case, gender, and number (CGN) has already been sketched within this framework. What remains is an account of the strong, weak, and mixed adjective (S, W, and W-MX) declensions as they relate to subtypes of determiners (I, II, and III).

Sections 3.2 through 3.5 examine a series of 'functional' proposals, all versions of the idea (presented in section 3.1) that the relationship between determiner subtypes and adjective declensions follows from a general principle requiring characteristic--unambiguous and nonredundant--exponents of the morphosyntactic categories CGN. I am unable to concoct any adequate formulation of this proposal.

Section 3.6 points out that such a constraint would be both transderivational and (in part) phonological, therefore not available in GPSG rules in any case. However, it would be expressible in a surface filter, rather than in a rule of syntax, assuming that surface filters apply to morphophonological representations. Even this last (GPSG-acceptable) treatment, I argue, is inadequate. Syntactic rules must relate determiner subtypes and adjective declensions, and rules of allomorphy that are adequate for German (sketched in section 3.7) do not refer to functional notions like ambiguity and redundancy and do not even have to refer to the phonological form of endings.

3.0. Property values and conventional references to them

In the interests of making it possible to formulate at least a few rules explicitly, I digress here on formal matters. Some readers might want to skip to the main body of the exposition in section 3.1.

In the remainder of this paper, for the sake of brevity and clarity I will use 'NOM', 'ACC', 'GEN', and 'DAT' to refer to the four cases of German; 'FEM', 'MASC', and 'NEUT' to refer to the three genders; 'SG' and 'PL' to refer to the two numbers; 'I', 'II', and 'III' to refer to the three determiner classes; and 'S', 'S-ES', 'S-ES-Z', 'W', and 'W-MX' to refer to the declension classes of adjectives and determiners. These are all to be understood as standing for property values, some of which were presented as simple in section 2 but are in fact best treated as complex.

For instance, I assume (following Bierwisch 1967) that the attribute CASE takes as its value a set of two properties, with bivalent attributes OBL (for the oblique cases, genitive and dative, versus the direct cases, nominative and accusative) and GOV (for the necessarily governed, or object, cases, accusative and dative, versus the ungoverned, or subject, cases, nominative and genitive). A reference to 'DAT' is then a reference to {OBL:+, GOV:+} as a value for CASE. I also assume (again following Bierwisch) that the attribute GEND takes as its value a set of two properties, with bivalent attributes F (for the feminine gender as against the masculine and neuter) and M (for the masculine gender as against the feminine and neuter). A reference to 'FEM' is then a reference to {F:+, M:-} as a value for GEND. And of course, 'SG' and 'PL' are references to - and +, respectively, as values for NUM.

Similar decompositions are needed for the properties of determiner class and adjective/determiner declension. Without defending these choices, I enumerate the properties I will be referring to below. The bivalent attribute INDC separates indeclinables (in particular, Class I determiners) from declinable modifiers (Class II and III determiners, and all adjectives). The bivalent attribute EIN separates the 'ein words' (the Class III determiners) from the 'der words' (the Class II determiners). The attribute DECL takes as its value a set of two properties, with bivalent attributes WK (for the weak and mixed declensions versus the strong declension) and MX (for the mixed versus the weak declension). As a result of these decisions, a reference to 'W-MX' is a reference to {WK:+, MX:+} as a value for DECL, and a reference to 'III' is a reference to - as a value for INDC in combination with + as a value for EIN.

3.1. The characteristic-exponent proposal

Faced with the complex details of agreement in German prenominal adjectives, some linguists--and language teachers--have sought a functional account of the facts. In particular, it has repeatedly been suggested that what lies behind the principles of adjective declension is the general condition that each CGN combination should have its own characteristic exponent at some point within an NP. On this proposal, the function of inflection (whether of a noun, a determiner, or an adjective) is to convey information about the morphosyntactic categories of the NP (cf. Durrell 1979: 71f.), and in the ideal case this information is conveyed both unambiguously and nonredundantly.

The characteristic-exponent proposal is sometimes presented to language learners via useful hints about how to remember the details of the adjective declensions, as in the following passages from an outline grammar of German, Eltzner and Radenhausen (1930):

> Weak Declension of Adjectives.--When an adjective is preceded by a der word, the case endings of the der word shows the gender, number, and case of the noun modified. The adjective, therefore, does not repeat these endings; it takes only the endings -e or -en. (p. 22)
>
> [Mixed Declension of Adjectives] When an adjective follows an ein word which lacks a case ending, the adjective supplies the ending...When the ein word has the characteristic case ending, the adjective has the weak ending... (p. 23)

The key word in the first quotation is therefore; adjectives, it is implicitly claimed, have distinctive endings only when these are not redundant expressions of CGN. Thus, klein 'little' in der kleine Mann 'the little man' takes the nondescript form kleine because the determiner der already indicates the CGN values NOM SG MASC. The key word in the second quotation is supplies; NPs, it is implicitly claimed, must have unambiguous indications of their CGN, and if these are not supplied by the determiner, they must be supplied by the adjective. Thus, klein in ein kleiner Mann 'a little man' has the strong form kleiner because this indicates the NOM (vs. ACC) and MASC (vs. NEUT) values not unambiguously supplied by the determiner ein.

3.2. The unadorned proposal

There are a number of complexities in turning these useful hints into a putative rule in the grammar of German. One was introduced in section 1.5 above: Head nouns bear (some) marks of case and number, and so can contribute something to the pool of CGN marks within an NP. Durrell (1979: 83) points out that noun forms can 'resolve ambiguities in the paradigm of the definite article', in cases like der Beamte 'the official' (NOM SG) vs. der Beamten 'the officials' (GEN PL) and die Fremde 'the (female) stranger' (NOM/ACC SG) vs. die Fremden 'the strangers' (NOM/ACC PL). We must decide whether a functionally based rule treats the entire NP 'as a morphological prime' (Durrell 1979: 82) or whether it takes only determiners and adjectives into account (as seems to be suggested in the Eltzner and Radenhausen quotations above).

Putting this issue aside for a moment, I observe that the simplest formulation of the characteristic-exponent proposal, (I) below, is obviously wrong.

(I) (a) A German NP *A* with CGN values *I* must contain inflectional material *M* making *A* unambiguously an exponent of *I*; that is, *M* must be phonologically distinct from the inflectional material in any NP with CGN values different from *I*.

(b) In addition, *M* must be a nonredundant exponent of *I*; that is, removing any of the inflected words in *A* must yield an *A'* that is phonologically identical to an NP with CGN values different from *I*.

A great many German NPs are ambiguous in the sense of (Ia), and some unambiguous NPs are redundant in the sense of (Ib). The NP *Frauen* 'women', for instance, is completely ambiguous as to its case, being either NOM, ACC, GEN, or DAT. And the NP *den Buechern* 'the books' (DAT) is unambiguous but redundant, since removing the determiner *den* yields an NP, *Buechern*, that is unambiguously DAT PL (its plurality indicated by umlaut and the suffix -*er*, its dative case indicated by the final suffix -*n*).

3.3. *First restriction*

Perhaps the conditions affect not all NPs, but only those with prenominal adjectives:

(II) (a) A German NP *A* having CGN values *I* and containing a prenominal adjective must contain inflectional material *M* making *A* unambiguously an exponent of *I*.

(b) In addition, *M* must be a nonredundant exponent of *I*.

But (II) will not do either. NPs like *das grosse Buch* 'the large book' (NOM or ACC), *die kluge Frau* 'the wise woman' (NOM or ACC), and *einer kluger Frau* 'a wise woman' (GEN or DAT) are all ambiguous as to case, and no inflectional affix carries the information that *des grossen Buch(e)s* 'the large book' (GEN) is NEUT rather than MASC, or that *einem grossen Tisch* 'a large table' (DAT) is MASC rather than NEUT. Moreover, the NPs *grossen Buechern* 'large books' and *den grossen Buechern* 'the large books' are redundant, since removing either the determiner *den* or the adjective *grossen* yields the unambiguous *Buechern* again.

3.4. *Two further restrictions*

Two amendments now suggest themselves, one for (IIa), the other for (IIb). The counterexamples I advanced to (IIa) involved CGN distinctions that are *never* indicated by inflectional material within an NP: NOM and ACC are always identical in form in the NEUT and FEM SG and throughout the PL; GEN and DAT are always identical in the FEM SG; and MASC and NEUT are always identical in the GEN and DAT SG. Consequently, one might revise (IIa) along the following lines:

(III) (a) A German NP *A* having CGN values *I* and containing a prenominal adjective must contain inflectional material *M* sufficient to make *A* phonologically distinct from any NP having CGN values *I'*, where *I* and *I'* are different CGN values that receive phonologically distinct exponents for at least one form class of German.

For (Ib) and (IIb), my counterexamples involved CGN values that were unambiguously indicated by noun inflection: *Buechern* can only be DAT PL. Consequently, one might revise (IIb) so as to focus on prenominal material only, along the following (somewhat hazy) lines:

(III) (b) In addition, inflectional affixes on a prenominal adjective must not supply information about *I* already supplied by those on a determiner.

One might have thought that by making the characteristic-exponent conditions so astoundingly particular--by now, they are generalizations over very small finite collections of relevant data--I would have succeeded in protecting them from counterexamples. But no. The NP *den grossen Flicken* 'the large patch(es)', which is either ACC SG or DAT PL (the MASC noun *Flicken* 'patch' being phonologically unaffected by shifts in case and number), serves as a counterexample to (IIIa). And the NP *eine kluge Frau* 'a wise woman' (NOM/ACC SG FEM) serves as a counterexample to (IIIb), because both the determiner *eine* and the mixed-declension adjective *kluge* distinguish the NOM/ACC SG FEM from all other CGN values: the indefinite article *eine* has no PL forms, and it has the ending -*e* in the SG only in the NOM/ACC FEM (see Table 4 in section 1.3); and the mixed declension of adjectives has -*e* only in the NOM/ACC FEM SG (see Table 3 in section 1.1).

3.5. *A final round of restrictions*

I believe that the characteristic-exponent proposal cannot be made to cover the facts for all three declensions of German. We might, however, lower our sights still further and try to describe only the mixed declension, taking the other two declensions as given. This restriction won't help (IIIb), however, since the counterexample to it in the previous paragraph involved the mixed declension. So we abandon the fight against redundancy and constrict the field of battle against ambiguity by two-thirds:

(IV) A German NP A with CGN values I, a determiner of Class III, and a prenominal adjective must contain inflectional material M sufficient to make A phonologically distinct from any NP having CGN values I', where I and I' are different CGN values that receive phonologically distinct exponents for at least one form class of German.

Incredibly enough, even though (IV) has a tiny domain, there is at least one type of counterexample, illustrated by the MASC NP meinen grossen Flicken 'my large patch(es)', which is ambiguous between ACC SG and DAT PL. I conclude that further contention is pointless, and declare the characteristic-exponent proposal vanquished.

Undoubtedly, the language exhibits some tendency towards characteristic exponents, and it is utterly reasonable that it should do so (otherwise, there would be no function for the inflectional apparatus of adjectives to perform and it should wither away over the generations--as, in fact, in some dialects of German it has). But there is no rule enforcing characteristic exponents.

3.6. The status of these proposals

What if one of these proposals had turned out actually to describe the facts of German? They are all generalizations about the surface forms of NPs in German. And powerful generalizations at that, for they are transderivational in character (they require that different paradigms be compared, rather than that one structure, or even one derivation for that structure, be examined) and also refer to phonology, morphology, and syntax all at once (they are sensitive to the phonological identity of inflectional affixes within a particular syntactic constituent type).

On both grounds, they could not possibly be encoded in GPSG (meta)rules: clearly, neither derivational nor transderivational reference is possible in the framework I sketched in section 2; and, as Pullum and Zwicky (1984) point out, reference to phonology is also out of the range of a GPSG syntax. Even in a transformational framework they would be extraordinary: transderivational constraints have not found wide acceptance in such frameworks, and it was proposed as long ago as Zwicky (1969) that reference to phonology in transformational rules should be prohibited. That is, there are good reasons for supposing that even if a principle like (I)-(IV) had turned out to be correct, it would not function as a condition on the application of a syntactic rule.

If such a principle is to be any sort of grammatical generalization, it must be a _surface filter_, a condition on the surface form of NPs in German. Such an analysis would be possible in a transformational framework, but at first glance it would appear to be inconsistent with GPSG. Certainly, the GPSG program does not countenance negative conditions, that is to say filters, in addition to its positive conditions, that is to say its phrase structure rules; the only negative statement about syntactic structure in GPSG is the (universal) final clause of a recursive definition: nothing is a phrase structure rule except by virtue of this definition. In any case, the transderivational and phonological nature of (III) would eliminate it as a candidate for a filter even in an extension of GPSG that embraced negative conditions.

However, there are arguments (alluded to in Zwicky (1983), developed in Zwicky and Pullum (forthcoming)) that surface filters apply not to _syntactic_ surface structure, but rather to a level of _morphophonological_ representation, namely the output of rules of allomorphy. As a theory of syntax, GPSG says nothing directly about phonology, although it has some indirect consequences for phonological theory (see Pullum and Zwicky (1984) on the Principle of Superficial Constraints in Phonology). Surface filters referring to phonology and morphology, even with transderivational power, are not ruled out in principle. A generalization like (I)-(IV) might then have a natural place as a surface filter.

But even this is not to be. Consider why surface filters are posited in the first place. In Perlmutter's original presentation (1971), a surface filter eliminates a configuration arising from the operation of several different rules (either separately or in interaction with one another). The rules are then permitted to apply without restriction, and the filter applies to the outputs resulting from the full set of rules.

In the German case we have been examining, the rules in question would include those distributing the values of adjective declension (S, W, and W-MX), those distributing the values of determiner class (I, II, and III), and allomorphy rules spelling out combinations of CGN values with declension class as particular endings. In a filter analysis, the declension class values would be freely distributed with respect to the determiner class values; endings would be freely distributed as exponents of the CGN/declension values; and (finally) principles like (I)-(IV) would act to eliminate distributions of endings which were either ambiguous or redundant.

Unfortunately, such principles just can't do enough work. There are many ways of achieving an unambiguous, nonredundant distribution of endings, given the available stock of them. A principle like (I)-(IV) cannot predict the particular distribution of endings German exhibits; these must, at least in part, be listed as the exponents of particular CGN/declension values.

3.7. An adequate description of the mixed declension

Although (IV) is inadequate as a generalization about German NPs and could not predict the actual endings of German NPs even if it had happened to be adequate, it can serve as the germ for an analysis of the allomorphy side of the phenomenon.

The key is to treat the weak and strong declensions as truly 'given' when the mixed declension allomorphs are being realized. I will assume that allomorphy rules say (a) for adjectives, what the phonological realization is for any CGN/declension combination if the value of DECL is S or W; and (b) for determiners, what the phonological realization is for any CGN/declension combination. Some of these allomorphy rules are generalizations, not mere spellings-out. One says that the ACC SG MASC (S or W) is -en, and another that the default for the ACC SG is to be identical to the NOM SG. One says that the NOM SG W ending is -e, and another that the default for W is -en. And so on.

What remains is to account for the mixed-declension endings on the basis of the endings in the other two declensions. The mixed declension of adjectives (Table 3) differs from the weak declension (Table 2) in only two respects, NOM SG MASC and NOM/ACC SG NEUT, which have the endings -er and -es, respectively, both drawn from the strong declension (Table 1). These are all the places, and the only places, where Class III determiners (which condition the mixed declension) have zero endings. The following generalization, which mentions neither ambiguity nor redundancy, is then true for German:

(V) The ending of an adjective in the mixed declension is chosen from the strong paradigm if the preceding determiner has a zero ending, otherwise from the weak paradigm.

Principle (V) actually predicts what the mixed-declension endings are, and it does so correctly, but it is still not a trouble-free allomorphy rule. It refers to the makeup of a word adjacent to the one whose inflectional apparatus is being described, and it refers to (phonological) zero. The reference to the internal composition of other words is, I believe, unparalleled in a rule of allomorphy.

However, we can take a clue from the description above of where the weak ending -e occurs, and reformulate (V) entirely in terms of word-internal morphological properties:

(VI) The ending of an adjective in the mixed declension is chosen from the strong paradigm in the non-FEM* NOM SG, otherwise from the weak paradigm.

Rule (VI) covers the NOM SG MASC and NOM SG NEUT occurrences of -e directly. It covers the ACC SG NEUT by virtue of the assumption (above) that the default for the ACC SG ending is to be identical to the NOM SG, and this instruction is not countermanded by any other statement. It does not cover the ACC SG MASC, because this ending is explicitly specified (above, again) as -en.

The allomorphy rule (VI) accounts correctly for the forms of the mixed paradigm on the basis of those in the strong and weak paradigms, and it does so without extravagant theoretical moves. There are many details to be worked out; in particular, the mechanisms of default setting need attention, as do those that have the effect of setting one ending identical to another. But so long as the declension values S, W, and W-MX are distributed correctly in phrase structures, allomorphy rules along the lines of (VI) can describe the morphological exponents of CGN values.

A final note: A reasonably explicit formulation of (VI) can be constructed, given the assumptions of section 3.0. What (VI) says is that something with a category C not distinct from {CAT:A, BAR:0, HEAD:{AGR:{CASE:{OBL:-, GOV:-}, GEND:{F:-}, NUM:-}}, DECL:{WK:+, MX:+}} takes endings identical to those for category C', where C' is derived from C by changing the value of DECL to {WK:-, MX:-}. The rule does not have to say that the weak paradigm is the default case; this is an automatic consequence of treating the mixed declension as a subtype of the weak declension, a decision made back in section 1.1 and formalized via the property WK:+ in section 3.0.

4. Analyzing the German adjective declensions in GPSG

Two potential mechanisms for describing the relationship between determiner subtypes and adjective declensions were presented in section 2: subcategorization of adjectives with respect to determiners (which I consider in section 4.1) and property agreement via the CAP and HFC (which I consider in section 4.2). The first is unsuitable for the case in hand, and the second turns out not to be available.

In section 4.3 I present an analysis in which this aspect of German adjective 'agreement' is in fact treated as government. The analysis is built around two principles in the metagrammar for German, Declension Government and

Declension Inheritance, though a number of other principles and default settings must be made explicit if the workings of these two are to be understood. The analysis also gives rise to some general questions about the description of government in GPSG, briefly surveyed in section 4.4.

4.1. Subcategorization

I turn now to the question of how to describe the implicational relationship between the determiner Class values I, II, and III and the adjective Declension values S, W, and W-MX. In line with the discussion of the preceding section, I take describing this relationship to be the only aspect of adjective declension in German that syntactic rules are responsible for; everything else is a matter of morphology and rules of allomorphy.

One mechanism GPSG provides for describing relationships between properties of nodes is subcategorization (by rule index; see section 2.8). A lexical category introduced in a rule is subcategorized with respect to the other constituents introduced by the same rule. The rule NP ---> Det Nom introduces the lexical category Det. Accordingly, determiners can be lexically marked as to whether they occur with Nom as their only sister under NP (there might be other rules introducing Det as a daughter of NP).

For our purposes, the subcategorization mechanism determines things in the wrong direction: the category Det is subcategorized by Nom, rather than the other way around. The property determined by the subcategorization mechanism is the wrong one: occurrence with Nom in general, rather than occurrence with Nom of the subtype S, W, or W-MX. And the node subcategorizing Det is the wrong one: Nom rather than its daughter A. Even if we wanted to have Det subcategorized by A, rather than the other way around, we would have to deal with the fact that Det and A are not sister nodes, hence cannot affect one another directly in GPSG.

The only subcategorization analysis that I can construct has Declension S, Declension W, and Declension W-MX as properties of A which must be duplicated as properties of the Nom node above A; then these properties subcategorize Det. Three things are peculiar about this analysis. First, it must treat the strong/weak/mixed distinction as lexically associated with adjectives; but the distinction is not lexical at all. Second, some parochial rule must insure that these properties of A are duplicated as properties of Nom; neither the HFC nor the FFP can be called on, since the properties in question surely are not on the UHFL or the UFPL. Third, the rule introducing Nom and Det as sisters must explicitly mention these properties of Nom, if subcategorization is to be invoked. This analysis can be made to work, but it is eminently unsuitable.

4.2. Agreement

Another mechanism that might be appealed to is the one provided by the CAP and the HFC working together. However, this is completely unavailable, because the CAP and HFC can apply only to properties on the UHPL, and such parochial properties as Class II and Declension S are certainly not on that list, given that they seem to have no semantic correlates at all.

In fact, the PAR of section 2.6 prohibits even language-particular (meta)rules requiring property agreement, so that we are not free to construct an agreement account specifically for German.

(The use of the HFC would be odd even if the properties in question were on the UHPL. The Head Feature Convention would function to distribute within phrases properties that were not realized morphologically on the heads of those phrases; a head noun doesn't show any sort of morphological indication of the class of its determiner. Cooper (to appear) argues that the HFC should not be permitted to apply to such 'silent features'.)

Suppose we abandoned the requirement that properties figuring in the CAP be on the UHPL. We would still be unable to use the CAP to ensure that the declension properties are correctly distributed in German NPs. If the CAP is to say that 'the form of a functor depends on properties of its argument expression' (Bach 1983: 70), as GP clearly intend it to,[7] then the determination of declension class runs in the wrong direction--Det is certainly the functor, Nom the argument expression, but the form of Nom depends on properties of Det--and the CAP is inapplicable.

Things are no better if, noting that the determiner class properties are associated with specific lexical items, we attempt to treat properties like Class II as foot properties rather than head properties, and so appeal to the FFP. Class II is no more likely to be on the UFPL than on the UHPL, and even if we gave up the UFPL, the FFP would only require that Class II on Det be duplicated as Class II on NP; it would not ensure that Class II, or some reflex of it, appeared on A.

4.3. Government

The analysis I opt for here is built on an observation made in section 1.4: Determiners of Classes I, II, and III require declension S, declension W, and declension W-MX, respectively, in their associated adjectives, but belong themselves to an Indeclinable set, declension S-ES, and declension S-ES-Z, respectively.

This is no sort of agreement.[8] Rather, a lexical class cleavage in the determiners is projected onto the adjectives as differences in inflection. 'The form of an argument depends on properties of the functor', as Bach (1983: 70) has it in his delineation of government. To put it yet another way, in agreement the head of a construction determines the form of a modifier, while in government a modifier determines the form of the head (see Zwicky 1984: sec. 2 for further discussion). The part of German adjective agreement that involves the determination of declension class is not agreement at all, but rather government.

What the syntax of German must say is that determiners of Class I impose the S declension on a following adjective, that determiners of Class II impose the W declension, and that determiners of Class III impose the W-MX declension. In a GPSG framework, this cannot be done in one step, since Det and A are not coconstituents. Recall the discussion in section 2.5: Det and Nom are coconstituents under NP, Nom branches into AP and N, AP branches into A', and A' branches into A. Two different principles are called for, one imposing properties on Nom by virtue of properties belonging to Det, the other propagating these properties 'down' from Nom, eventually to A. I will call these principles Declension Government (DG) and Declension Inheritance (DI), respectively.

Both of these principles belong to the metagrammar. DG acts as a rider on the branching of NP into Det plus Nom, DI as a rider on any branching of a category X into some set of categories, one of which is AP, A', or A--that is, one of which has the property CAT:A. Formulating the latter is straightforward:

> Declension Inheritance: If category X has a daughter category Y with the property CAT:A, then X and Y must have identical values for the attribute DECL.

DI is reminiscent of the HFC; both require identity of certain properties between AP and A', and between A' and A. But it could not be collapsed with the HFC even if properties with the attribute DECL were on the UHPL, for the HFC does not require property identity between Nom and its modifier daughter AP, and the DI does.

Formulating DG is a trickier business, and requires some use of the formalism developed in section 3.0, because the exact shape DG takes will depend on how the default values for WK and MX in DECL are chosen; DG need mention only properties of Nom that have nondefault values, all remaining properties being filled in by default. For WK, at least, there is fairly clear evidence about the default. Recall from section 1.1 that the strong declension of adjectives is used both with invariable determiners like zwei 'two' and also with a zero determiner, as in the mass NP braunes Bier 'brown beer' (NOM/ACC) and the plural NP brauner Buecher 'of brown books' (GEN). The natural analysis for such NPs is that they have no Det, not that they have determiners whose phonological realizations are null; certainly the null-determiner analysis would require justification. If there is a branching of NP into only one daughter, Nom, then there is no Det to govern an adjective declension in this construction, and the adjective declension that appears there must be the default. It follows that WK has the default value –. As for MX, I will assume that it too has a – default:

> Defaults for DECL: In any category with the property CAT:A, the default value for WK and MX in DECL is –.

The task of DG is then to say when Nom has the properties WK:+ and MX:+. The first property is predictable from the fact that Det is declinable, the second from the fact that Det is an ein word:

> Declension Government: In a branching of NP into Det and Nom, if Det has the property INDC:– then Nom has the property WK:+ in its value for DECL; and if Det has the property EIN:+ then Nom has the property MX:+ in its value for DECL.

With these formulations of DG and DI, the main part of my description of the German adjective declensions is finished. There are still some details worth discussing, having to do with the fact that German adjectives are sometimes declined, sometimes indeclinable.

The large generalization about this phenomenon is that adjectives are declined only when they are prenominal; I will disregard further details here. We need to describe the contrast between Die Frau ist klug 'The woman is wise', with the undeclined adjective form klug, and die kluge Frau 'the wise woman', with a declined form. The attribute in question is INDC, which I will say has the default value + ('indeclinable') for adjectives. This default is overridden within a prenominal AP, that is, within an AP that is the daughter of Nom.

An additional wrinkle comes in the fact that there are some reasons (not the least being their inflectional paradigms) for grouping the open classes of adjectives and nouns together with the closed classes of personal pronouns and determiners, at least in German; and the default value of INDC for all of these classes except the adjectives is certainly –. The natural property for these four groups of lexical items to share is the property N:+, in the system that GP provide for the analysis of the major word-class properties N (= {N:+, V:–}), A (= {N:+, V:+}), V (= {N:–, V:+}), and P (= {N:-, V:-}). Putting all of these observations about declinability together, we have the two following principles:

Prenominal Adjectives: AP as a daughter of Nom has the property INDC:-.

Defaults for INDC: The default value for INDC is + for categories with the property CAT:A. The default value for INDC is - for categories with the property CAT:{N:+}.

The value settings run through three levels here: INDC for an AP daughter of Nom has the value -, overriding the default + value for categories with the property CAT:A, which in turn overrides the default - value for categories with the property CAT:{N:+}.

Two important issues having to do with the attribute INDC remain. The first is that values of INDC and DECL are distributed independently of one another by the principles above, but they are of course not independent. Indeed, as things stand free instantiation would allow both the appearance of DECL in a predicate adjective, where it would get a default value of {WK:-, MX:-}, and a default assignment of INDC:+ to the same predicate adjective; the latter ought to prevent the former. The second, closely related, problem is that the Prenominal Adjectives principle above assigns INDC a - value only at the AP level, but the place where INDC does its real work is at the A level, where it determines whether or not rules of allomorphy realize properties of words as inflectional affixes; free instantiation of INDC should be prevented from assigning INDC:+ to an A dominated by an AP with the property INDC:-. Another inheritance principle could be stated, but it would solve only the second problem. Both problems can be solved by preventing free instantiation--in the first case, of DECL (with any value) in a predicate adjective having the property INDC:+; in the second, of INDC (with the value +) in a prenominal adjective having any value for DECL. The following principle does the trick:

Declinability: A category has the property INDC:+ if and only if it has no property with the attribute DECL.

This principle connects a property determining the applicability of some set of morphological rules with a property that (in effect) picks out the applicable rule. The connection is obviously not a matter of German grammar, but a universal generalization about systems of properties.

This completes the sketch of the syntactic side of adjective agreement in German. Two universal metagrammatical principles, the HFC and CAP, require that the case, gender, and number properties of N or NP be duplicated on a prenominal adjective. Two principles of the metagrammar for

German, DG and DI, in concert with default settings and two principles involving INDC (one parochial, one universal), ensure that the adjective has the declension property appropriate to its context.

A syntactic analysis along these lines ought to be constructible regardless of the sort of morphological analysis it is destined to be combined with. The details will vary with the morphological framework, of course, and there is no denying that my analysis is tailored to a specific view of inflectional morphology. In this view, inflectional morphology describes the way in which (bundles of) morphosyntactic properties of words are realized as affixes (or morphological processes, which I have not dealt with here). The primary descriptive tool is the rule of allomorphy, which either assigns phonological content to the properties or refers the assignment to another combination of properties (as when the assignment for the ACC SG is referred to that for the NOM SG, or when the assignment for the mixed declension is referred to that for the strong declension under certain conditions). Like my syntactic analysis, this approach to inflectional morphology relies heavily on principles (some of them rather complex) giving default assignments, with competition between principles resolved in favor of the more specific principle (as when the assignment for the ACC SG MASC overrides the assignment for the ACC SG).[9]

4.4. On the analysis of government in GPSG

I move now to wider issues concerning government and its analysis in a GPSG framework, which I will approach by observing some differences in the phenomena to be analyzed.

GPSG permits the description of two different sorts of phenomena falling under the traditional heading of government: what I will call 'vertical government' and 'horizontal government'. In vertical government a category has a property by virtue of appearing as a daughter of some specified category. In horizontal government, the familiar type, a category has a property by virtue of appearing as a sister of some specified category.

Vertical government can be illustrated by English prenominal possessives like *this evening's* in *this evening's events*. These can be analyzed as NP determiners, with the CASE:GEN property supplied in the rule licensing the branching of Det into a lone NP. Thus, the NP has this property by virtue of appearing as a daughter of Det, rather than S, VP, or PP. A similar analysis might be entertained for nominative NPs in English, if it is assumed that CASE:ACC is the default assignment for CASE, so that it is the task of some syntactic rule(s) of English to say where CASE:NOM occurs. On these assumptions, the CASE:NOM property would be supplied in the rule licensing the branching of S into NP and VP. The NP would have this property by virtue of appearing as the daughter of S, rather than Det, VP, or PP.

A horizontal-government treatment of the nominative case is also available, of course, and it is indistinguishable from the vertical-government treatment unless either (a) NP and VP can be sisters under some category other than S, or (b) NP and some category other than VP can be sisters under S (or NP can appear as the sole daughter of S). Failing that, both analyses simply say that S can branch into a VP and an NP with the property CASE:NOM. If (a) or (b) holds, then the possibility arises that vertical and horizontal government can be distinguished, and also that an instance of government should be described not in a rule but in the metagrammar, as a generalization across all rules of a certain type. That is, it might be that NP has the property CASE:NOM in any rule licensing it as a daughter of S. Or that NP has this property in any rule licensing it as a sister of VP.

Note, furthermore, that in standard examples of horizontal government the governing category is a lexical category. Verbs and prepositions, for instance, govern particular cases of their object NPs. If the relationship between a complementizer and the S it combines with is viewed as government of the S by the complementizer (so that that governs a finite S, for an infinitive S, wh-words a slashed finite S, etc.), then this too is horizontal government with a lexical category serving as the governor. So there is some question as to whether a horizontal-government analysis of nominative case should be available, since the governing category would be the phrasal category VP.

In some instances of horizontal government, the governing category is not only lexical, but also at least in part arbitrary. In languages in which verbs or prepositions can govern several different cases, for instance, it is typical that one cannot predict, on the basis of their syntactic or semantic properties, exactly which items govern a nondefault case; the class of governors is partly arbitrary. This is certainly true for the German verbs and prepositions governing the DAT or GEN rather than the default ACC. And it is true for the German determiners governing declension properties, as I observed in section 1.2.

An important difference between case government and declension government in German is that in the former the determined properties (with the attribute CASE) are on the UHPL, but in the latter the determined properties (with the attribute DECL) are parochial. As one result of this difference, the determined properties in the former example (but not the latter) participate in agreement via the HFC and CAP.

Another result of this difference is that the former phenomena, but not the latter, have an alternative analysis that is not, speaking intuitively at least, government. A horizontal-government analysis of DAT and GEN case-marking in German says: In a branching of VP into V and NP (and possibly other categories), if V has the property SUBCLASS:X then NP has the property CASE:DAT, and if V has the property SUBCLASS:Y then NP has the property CASE:GEN. In the alternative, a rule schema permits the NP in such a branching to occur with any one of the properties CASE:ACC, CASE:DAT, or CASE:GEN, and V is then subcategorized according to these properties. The alternative makes the analysis of case-marking in German entirely parallel to the subcategorization of English verbs according to whether they occur with various types of objects (one NP, two NPs, one NP plus a PP in to, one NP plus a PP in for, etc.).

Let me now pull some of these analytic threads together. What lies behind the preceding discussion is a concern that the theory of grammar should constrain government in much the same way that it constrains agreement. Can horizontal government be restricted to instances with a lexical category as governor? (If so, then nominative case-marking in English must be vertical government.) Or to instances with parochial governed categories? (If so, then the subcategorization analysis is the only one available for object case-marking in German, in which case the phenomenon is not treated by the grammar as government in a strict sense.) Can metarules for vertical government be prohibited? (If so, then vertical government disappears as a substantive notion in GPSG, since it is then merely the appearance of some property on a daughter category specified by a rule and is indistinguishable from a simple instance of horizontal government.)

I do not know what the answers to these questions are, though I am inclined to suppose that they are all positive. Certainly the questions are worth further investigation.

Notes

*Grateful thanks to the Center for the Study of Language and Information, Stanford University, whose financial support enabled me to complete this paper. And to Geoffrey K. Pullum for his comments on earlier drafts of section 2.

1. There are principles governing which endings are identical to which others. For instance: for all gender/number combinations except masculine singular, the accusative is identical to the nominative. See section 3.7 for further development of this idea.

2. The definite article has idiosyncratic allomorphy in the nominative and accusative forms: wherever des would be expected in these forms, das occurs instead, and wherever de would be expected in these forms, die occurs instead. (As it happens, des does occur where expected in the genitive.) These facts about allomorphy do not affect the syntactic generalizations to be made.

3. The variation between -es and -s in the genitive singular of masculine and neuter nouns in Tables 6 and 8 is phonologically conditioned, and will not concern us here. Similarly, the variation between -e and zero in the dative singular of masculine and neuter nouns is phonologically and stylistically conditioned, and will not concern us here.

4. Listing the HFC, CAP, and FFP does not preclude the existence of other general principles governing the distribution of features in constituent structures. Indeed, Gazdar, Klein, Sag, and Pullum (1982) entertain an analysis of conjunction in which the feature CONJ, neither a head nor a foot feature, obeys its own (universal) principles of occurrence.

5. In Lapointe's scheme, there are no syntactic rules of agreement, hence no 'grammatical agreement' in the usual sense. Instead, agreement facts are supposed to fall out from a well-formedness condition on Logical Form plus the well-formedness condition on S-structures.

6. Note that in the system of property values adopted in section 3.0, the non-FEM genders constitute a natural class, namely the class of categories with the property GEND:{F:-}.

7. The version of the CAP that GP give is completely symmetrical and does not in itself reflect any logical directionality in the relationship between the determinans and determinatum in grammatical agreement.

8. Nor any sort of disagreement rule, which is what Lapointe suggests on the basis of a simplified set of paradigms.

9. This exploitation of a generalized Proper Inclusion Precedence, or 'elsewhere', condition on morphological rules it shares with lexical, or level-ordered, morphology (see Kiparsky 1982 and the references therein), with which it is not in principle inconsistent.

References

Anderson, Stephen D. 1982. Where's morphology? LingI 13.571-612.

Bach, Emmon. 1983. On the relationship between word-grammar and phrase-grammar. NLLT 1.1.65-89.

Barlow, Michael, Daniel Flickinger, and Ivan A. Sag (eds.). 1982. Developments in Generalized Phrase Structure Grammar. Bloomington: Indiana Univ. Linguistics Club.

Bierwisch, Manfred. 1967. Syntactic features in morphology: General problems of so-called pronominal inflection in German. To Honor Roman Jakobson I.239-70. The Hague: Mouton.

Cooper, Robin. To appear. Swedish and the Head Feature Convention. In L. Hellan and K. Koch Christensen (eds.), Topics in Scandinavian syntax. Dordrecht: D. Reidel.

Curme, George O. 1960. A grammar of the German language. 2nd ed. New York: Frederick Ungar.

Durrell, Martin. 1979. Some problems in the morphology of the German noun phrase. TPS 1979.66-88.

Eltzner, Dorothea and Paul Radenhausen. 1930. German. New York: Oxford Book Co.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan Sag. 1982. Coordinate structure and unbounded dependencies. In Barlow et al. 1982: 38-68.

Gazdar, Gerald and Geoffrey K. Pullum. 1982. Generalized Phrase Structure Grammar: A theoretical synopsis. Bloomington: Indiana Univ. Linguistics Club.

Keenan, Edward L. 1974. The functional principle: Generalizing the notion of 'subject of'. CLS 10.298-309.

Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In van der Hulst and Smith 1982: 131-75.

Lapointe, Steven G. 1983. A comparison of two recent theories of agreement. In Richardson et al. 1982: 122-34.

Perlmutter, David M. 1971. Deep and surface structure constraints in syntax. New York: Holt, Rinehart and Winston.

Pullum, Geoffrey K. and Arnold M. Zwicky. 1984. The syntax-phonology boundary and current syntactic theories. OSU WPL 29.105-16.

Richardson, John F., Mitchell Marks, and Amy Chukerman (eds.). 1983. Papers from the parassession on the interplay of phonology, morphology, and syntax. Chicago, Ill.: CLS.

van der Hulst, Harry and Norval Smith (eds.). 1982. The structure of phonological representations (Part I). Dordrecht: Foris.

Wasow, Thomas. 1978. On constraining the class of transformational languages. Synthese 39.81-104.

Zwicky, Arnold M. 1969. Phonological constraints in syntactic descriptions. Papers in Linguistics 1.3.411-63.

Zwicky, Arnold M. 1983. An expanded view of morphology in the syntax-phonology interface. PICL 13.198-208.

Zwicky, Arnold M. 1982. Heads. OSU WPL 29.50-69.

Zwicky, Arnold M. and Geoffrey K. Pullum. Forthcoming. The syntax-phonology interface. New York: Academic Press.

This is the version of August 1984.